STIC-Biotech/ChemLib 160809 LB

CRFE

From: Whiteman, Brian
Sent: Thursday, July 28, 2005 8:39 AM
To: 78544 STIC-Biotech/ChemLib
Subject: seq search

10/782,899 2/23/04 Fujimori et al.

nucleotides 1-192 of SEQ ID NO: 1, nucleotides 472-600 of SEQ ID NO: 1
1) the commercial databases, and the issued and published US application databases

Thank you,
Brian Whiteman
Remsen, 2D14
mail box 2C18
Patent Examiner - Art Unit 1635
United States Patent and Trademark Office
(571) 272-0764

1- 600 na
LB

****
STAFF USE ONLY
Searcher:______________
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed:____________
Searcher Prep/Rev. Time:___
Online Time:____________

****
Type of Search
NA#:______ AA#:______
Interference:____ SPDI:____
S/L:______ Oligomer:_____
Encode/Transl:___________
Structure#:______ Text:____
Inventor:____ Litigation:___

****
Vendors and cost where applicable
STN:________________
DIALOG:______________
QUESTEL/ORBIT:__________
LEXIS/NEXIS:___________
SEQUENCE SYSTEM:_______
WWW/Internet:__________
Other(Specify):__________

Date completed: 8-8-05
Searcher: Beverly e 2528
Terminal time: ______________
Elapsed time: ______________
CPU time: ______________
Total time: ______________
Number of Searches: ______________
Number of Databases: ______________

Search Site
______ STIC
______ CM-1
______ Pre-S

Type of Search
______ N.A. Sequence
______ A.A. Sequence
______ Structure
______ Bibliographic

Vendors
______ IG
______ STN
______ Dialog
______ APS
______ Geninfo
______ SDC
______ DARC/Questel
✓ Other CGN

PTO-1590 (9-90)

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 23:44:09 ; Search time 282.514 Seconds
                                           (without alignments)
                                           2959.924 Million cell updates/sec

Title:          US-10-782-899-1_COPY_472_600
Perfect score:  129
Sequence:       1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       7297361 seqs, 3241162794 residues

Total number of hits satisfying chosen parameters:      14594722

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
                18:  /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
                19:  /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
                20:  /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
                21:  /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
                22:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                23:  /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
                24:  /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
                25:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                26:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                    %
Result            Query
  No.    Score    Match Length DB  ID                      Description
 -----------------------------------------------------------------------
    1      129    100.0    600  9  US-09-816-391A-1         Sequence 1, Appli
    2      129    100.0    600 21  US-10-782-899-1          Sequence 1, Appli
c   3      129    100.0 2256646 19  US-10-470-565-1          Sequence 1, Appli
    4     29.4     22.8    885 22  US-10-617-320-306        Sequence 306, App
c   5     29.4     22.8   1377 19  US-10-687-677-1          Sequence 1, Appli
    6     29.4     22.8   3520 10  US-09-814-353-19115      Sequence 19115, A
c   7     29.2     22.6    634 18  US-10-332-859-15         Sequence 15, Appl
    8     29.2     22.6   1013 18  US-10-332-859-232        Sequence 232, App
    9     29.2     22.6   1836 18  US-10-332-859-17         Sequence 17, Appl
   10       29     22.5 344805 20  US-10-779-271-1          Sequence 1, Appli
   11       29     22.5 354592 22  US-10-737-082-70         Sequence 70, Appl
   12       29     22.5 354592 22  US-10-765-790-70         Sequence 70, Appl
   13     28.8     22.3    766 18  US-10-332-859-317        Sequence 317, App
c  14     28.6     22.2   1043 20  US-10-425-115-152385     Sequence 152385,
   15     28.4     22.0    936  9  US-09-903-410-31        Sequence 31, Appl
   16     28.4     22.0    936 13  US-10-027-805-31         Sequence 31, Appl
   17     28.4     22.0    936 14  US-10-027-804-31         Sequence 31, Appl
   18     28.4     22.0   2451  9  US-09-938-842A-64       Sequence 64, Appl
   19     28.4     22.0   2451 11  US-09-938-842A-64        Sequence 64, Appl
c  20     28.2     21.9    599 16  US-10-029-386-13097      Sequence 13097, A
c  21     28.2     21.9   3519 19  US-10-437-963-80607      Sequence 80607, A
c  22       28     21.7 165221 13  US-10-087-192-1015       Sequence 1015, Ap
c  23     27.8     21.6    483 18  US-10-424-599-128897     Sequence 128897,
c  24     27.8     21.6   2037 13  US-10-052-586-591        Sequence 591, App
c  25     27.8     21.6   2037 14  US-10-174-590-591        Sequence 591, App
c  26     27.8     21.6   2037 14  US-10-176-758-591        Sequence 591, App
c  27     27.8     21.6   2037 14  US-10-175-737-591        Sequence 591, App
c  28     27.8     21.6   2037 14  US-10-174-581-591        Sequence 591, App
c  29     27.8     21.6   2037 14  US-10-176-483-591        Sequence 591, App
c  30     27.8     21.6   2037 14  US-10-176-749-591        Sequence 591, App
c  31     27.8     21.6   2037 14  US-10-176-914-591        Sequence 591, App
c  32     27.8     21.6   2037 14  US-10-176-915-591        Sequence 591, App
c  33     27.8     21.6   2037 14  US-10-173-706-591        Sequence 591, App
c  34     27.8     21.6   2037 14  US-10-175-738-591        Sequence 591, App
c  35     27.8     21.6   2037 14  US-10-175-752-591        Sequence 591, App
c  36     27.8     21.6   2037 14  US-10-176-482-591        Sequence 591, App
c  37     27.8     21.6   2037 14  US-10-176-757-591        Sequence 591, App
c  38     27.8     21.6   2037 14  US-10-176-913-591        Sequence 591, App
c  39     27.8     21.6   2037 14  US-10-180-552-591        Sequence 591, App
c  40     27.8     21.6   2037 14  US-10-180-557-591        Sequence 591, App
c  41     27.8     21.6   2037 14  US-10-173-700-591        Sequence 591, App
c  42     27.8     21.6   2037 14  US-10-174-572-591        Sequence 591, App
c  43     27.8     21.6   2037 14  US-10-174-579-591        Sequence 591, App
c  44     27.8     21.6   2037 14  US-10-174-582-591        Sequence 591, App
c  45     27.8     21.6   2037 14  US-10-174-588-591        Sequence 591, App

                                 ALIGNMENTS

RESULT 1
US-09-816-391A-1
; Sequence 1, Application US/09816391A
; Patent No. US20020054865A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, Minoru
;  APPLICANT:  TANIGUCHI, Shunichiro
;  APPLICANT:  AMANO, Jun
;  APPLICANT:  YAZAWA, Kazuyuki
;  APPLICANT:  KANO, Yasunobu
;  APPLICANT:  NAKAMURA, Toshiyuki
;  APPLICANT:  SASAKI, Takayuki
;  TITLE OF INVENTION: Anaerobic bacterium as a drug for cancer gene therapy
;  FILE REFERENCE: 2001-WMC/01736
;  CURRENT APPLICATION NUMBER: US/09/816,391A
;  CURRENT FILING DATE:  2001-03-26
;  PRIOR APPLICATION NUMBER: JP 00/287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-09-816-391A-1

  Query Match             100.0%;  Score 129;  DB 9;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-37;
```

```
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 2
US-10-782-899-1
; Sequence 1, Application US/10782899
; Publication No. US20050025745A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, MINORU
;  APPLICANT:  TANIGUCHI, SHUNICHIRO
;  APPLICANT:  AMANO, JUN
;  APPLICANT:  YAZAWA, KAZUYUKI
;  APPLICANT:  KANO, YASUNOBU
;  APPLICANT:  NAKAMURA, TOSHIYUKI
;  APPLICANT:  SASAKI, TAKAYUKI
;  TITLE OF INVENTION: ANAEROBIC BACTERIUM AS A DRUG FOR CANCER GENE THERAPY
;  FILE REFERENCE: 671308-2001.1
;  CURRENT APPLICATION NUMBER: US/10/782,899
;  CURRENT FILING DATE:  2004-02-23
;  PRIOR APPLICATION NUMBER: 09/816,391
;  PRIOR FILING DATE: 2001-03-26
;  PRIOR APPLICATION NUMBER: JP 2000-287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-10-782-899-1

  Query Match             100.0%;  Score 129;  DB 21;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-37;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 3
US-10-470-565-1/c
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
;  APPLICANT: Societe des Produits Nestle S.A.
;  TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
;  FILE REFERENCE: 80290/WO
;  CURRENT APPLICATION NUMBER: US/10/470,565
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: EP 01102050.0
;  PRIOR FILING DATE: 2001-01-30
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 2256646
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
US-10-470-565-1

  Query Match             100.0%;  Score 129;  DB 19;  Length 2256646;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-36;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2241699 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 2241640

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2241639 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 2241580

Qy        121 AGTTCAGCG 129
              |||||||||
Db    2241579 AGTTCAGCG 2241571


RESULT 4
US-10-617-320-306
; Sequence 306, Application US/10617320
; Publication No. US20050136404A1
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/617,320
;             FILING DATE: 10-Jul-2003
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 306:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 885 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
```

```
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         ORIGINAL SOURCE:
;              ORGANISM: Streptococcus pneumoniae
;         FEATURE:
;              NAME/KEY:  misc_feature
;              LOCATION: (B) LOCATION 1...885
;         SEQUENCE DESCRIPTION: SEQ ID NO: 306:
US-10-617-320-306

  Query Match             22.8%;  Score 29.4;  DB 22;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 2.8;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
               ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db         604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy          92 CGTGTTGTCCAACCGTATTATTCCGGA 118
               |||| | |      ||   || | |||
Db         664 CGTGGTTTTGCCTGGTTGGATACGGGA 690

RESULT 5
US-10-687-677-1/c
; Sequence 1, Application US/10687677
; Publication No. US20040142419A1
; GENERAL INFORMATION:
;  APPLICANT: Guy, John
;  TITLE OF INVENTION: REDUCING CELLULAR DYSFUNCTION CAUSED BY MITOCHONDRIAL GENE
;  TITLE OF INVENTION:  MUTATIONS
;  FILE REFERENCE: 5853-324
;  CURRENT APPLICATION NUMBER: US/10/687,677
;  CURRENT FILING DATE:  2003-10-17
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-687-677-1

  Query Match             22.8%;  Score 29.4;  DB 19;  Length 1377;
  Best Local Similarity   63.4%;  Pred. No. 3.2;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          57 GAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
               | || | | ||  || | ||||||  ||| |||| |  | || || || || || ||  |
Db        1292 GTGAAGCTGGGCTTCATGTTGTTGATGTGGTGTGTCAGGCTGCCCCACTGTGTTGTTGTG 1233

Qy         117 GACTAGTTCAG 127
                ||  || |||
Db        1232 AACATGTACAG 1222

RESULT 6
US-09-814-353-19115
; Sequence 19115, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
```

```
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19115
;   LENGTH: 3520
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-19115

  Query Match             22.8%;  Score 29.4;  DB 10;  Length 3520;
  Best Local Similarity   54.1%;  Pred. No. 4.2;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy           4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
               || ||| ||  |      | |||      ||  | |    |   | |||   |  | |||
Db          27 CCCTCTACTGCTCATCTCTGCTTACCTTGTTTTTTAGTGGATTAATCCCTCACCTGGTGG 86

Qy          64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTC 114
                ||  ||  |||| |||||| ||||| || || || || ||   || || |
Db          87 GCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTCATAATCC 137

RESULT 7
US-10-332-859-15/c
; Sequence 15, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
;   LENGTH: 634
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-15

  Query Match             22.6%;  Score 29.2;  DB 18;  Length 634;
  Best Local Similarity   59.8%;  Pred. No. 2.9;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
               | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db         245 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 186

Qy         108 ATTATTCCGGACTAGTTCAGCG 129
               || | | |  ||| | |   ||
Db         185 ATCAATGCCTACTTGGTGGACG 164

RESULT 8
```

```
US-10-332-859-232
; Sequence 232, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 232
;   LENGTH: 1013
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-232

  Query Match             22.6%;  Score 29.2;  DB 18;  Length 1013;
  Best Local Similarity   59.8%;  Pred. No. 3.4;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        769 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 828

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        829 ATCAATGCCTACTTGGTGGACG 850


RESULT 9
US-10-332-859-17
; Sequence 17, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
;   LENGTH: 1836
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-17

  Query Match             22.6%;  Score 29.2;  DB 18;  Length 1836;
  Best Local Similarity   59.8%;  Pred. No. 4.1;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        813 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 872

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        873 ATCAATGCCTACTTGGTGGACG 894


RESULT 10
US-10-779-271-1
; Sequence 1, Application US/10779271
; Publication No. US20040220387A1
; GENERAL INFORMATION:
;  APPLICANT: AstraZeneca AB
;  TITLE OF INVENTION: Methods
;  FILE REFERENCE: ASZD-P02-251
;  CURRENT APPLICATION NUMBER: US/10/779,271
;  CURRENT FILING DATE:  2004-02-13
;  PRIOR APPLICATION NUMBER: 09/463,844
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: PCT/GB98/00259
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 9716162.4
;  PRIOR FILING DATE: 1997-08-97
;  PRIOR APPLICATION NUMBER: 60/535,986
;  PRIOR FILING DATE: 2004-01-12
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 344805
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-779-271-1

  Query Match             22.5%;  Score 29;  DB 20;  Length 344805;
  Best Local Similarity   55.4%;  Pred. No. 25;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db      47662 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 47721

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db      47722 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 47762


RESULT 11
US-10-737-082-70
; Sequence 70, Application US/10737082
; Publication No. US20050130170A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2032
;  CURRENT APPLICATION NUMBER: US/10/737,082
;  CURRENT FILING DATE:  2003-12-16
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70
;   LENGTH: 354592
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-737-082-70

  Query Match             22.5%;  Score 29;  DB 22;  Length 354592;
  Best Local Similarity   55.4%;  Pred. No. 25;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db      50053 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 50112

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db      50113 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 50153


RESULT 12
US-10-765-790-70
; Sequence 70, Application US/10765790
; Publication No. US20050130172A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70
;   LENGTH: 354592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-70

  Query Match             22.5%;  Score 29;  DB 22;  Length 354592;
  Best Local Similarity   55.4%;  Pred. No. 25;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db      50053 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 50112

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db      50113 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 50153


RESULT 13
US-10-332-859-317
; Sequence 317, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 317
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-317

  Query Match             22.3%;  Score 28.8;  DB 18;  Length 766;
  Best Local Similarity   58.5%;  Pred. No. 4.4;
  Matches   48;  Conservative    1;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  |:  |||| |||||| ||||| || || || || ||
Db        229 ATCCCTCACCTCCTGGGCGATGWGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 288

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        289 ATCAATGCCTACTTGGTGGACG 310


RESULT 14
US-10-425-115-152385/c
; Sequence 152385, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 152385
;   LENGTH: 1043
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_70558C.1
US-10-425-115-152385

  Query Match             22.2%;  Score 28.6;  DB 20;  Length 1043;
  Best Local Similarity   55.6%;  Pred. No. 5.7;
  Matches   55;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         19 GATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGT 78
              | | |||||   | ||  ||||   |  || | ||||  ||  ||||||| |||   ||
Db       1041 GCTGACTTCCCACCTTGATGACTCAAGGGAGCACGACGTAGTAGGTCGGGATCTGGGTGC 982

Qy         79 TGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGG 117
              |   | ||   | ||| ||  | || |      |||| |
Db        981 TCCTGAGCATCGGCGTCTTCACGAAACCCGAACTTCCCG 943


RESULT 15
US-09-903-410-31
; Sequence 31, Application US/09903410
; Patent No. US20020146799A1
; GENERAL INFORMATION:
;  APPLICANT: DIVERSA CORPORATION
;  APPLICANT:  ROBERTSON, Dan
;  APPLICANT:  MURPHY, Dennis
;  APPLICANT:  REID, John
;  APPLICANT:  MAFFIA, Anthony
;  APPLICANT:  LINK, Steven
```

```
;  APPLICANT:  SWANSON, Ronald
;  APPLICANT:  WARREN, Patrick
;  APPLICANT:  KOSMOTKA, Anna
;  TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
;  FILE REFERENCE: DIVER1180-2
;  CURRENT APPLICATION NUMBER: US/09/903,410
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: US 09/382,242
;  PRIOR FILING DATE: 1999-08-24
;  PRIOR APPLICATION NUMBER: US 08/602,359
;  PRIOR FILING DATE: 1996-02-16
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 936
;   TYPE: DNA
;   ORGANISM: Archaeoglobus fulgidus
US-09-903-410-31

  Query Match            22.0%;  Score 28.4;  DB 9;  Length 936;
  Best Local Similarity   56.4%;  Pred. No. 6.6;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              ||||||  |  | || |   | |   || ||| || | |||  | |||||     |||||
Db        718 GACCTTGAGAACCTACCTCCTGCGCTGATCATAACCGCCGAATACGACCCGCTGAGAGAT 777

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
              |   | |   |||| |    | ||||| || | |
Db        778 GAAGGAGAAGTTTTCGGGCAGATGCTGAGAAGAG 811


Search completed: August  6, 2005, 01:45:29
Job time : 287.014 secs
```

```
                      GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          August  5, 2005, 21:17:03 ; Search time 206.159 Seconds
                                             (without alignments)
                                             3704.166 Million cell updates/sec

Title:           US-10-782-899-1_COPY_472_600
Perfect score:   129
Sequence:        1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:      8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        N_Geneseq_16Dec04:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002as:*
                 7:  geneseqn2002bs:*
                 8:  geneseqn2003as:*
                 9:  geneseqn2003bs:*
                 10:  geneseqn2003cs:*
                 11:  geneseqn2003ds:*
                 12:  geneseqn2004as:*
                 13:  geneseqn2004bs:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                  %
Result           Query
  No.    Score   Match Length DB  ID                  Description
-----------------------------------------------------------------------------
    1      129   100.0    600  6  ABK52324            Abk52324 DNA encod
c   2      129   100.0 156638  6  ABQ81850            Abq81850 Bifidobac
    3     30.6    23.7  26309  4  AAS59535            Aas59535 Propionib
    4     30.6    23.7  26309  8  ACF64464            Acf64464 Propionib
    5     29.6    22.9   2453  10  ADE63675            Ade63675 Human gen
    6     29.4    22.8    885  13  ADR91671            Adr91671 Novel S.
c   7     29.4    22.8   1377  12  ADQ89190            Adq89190 Non-natur
    8     29.4    22.8   3520  5  ADL45225            Adl45225 Human ova
c   9     29.2    22.6    634  6  ABL01350            Abl01350 Murine ap
   10     29.2    22.6   1800  13  ADR89493            Adr89493 Apoptosis
   11     29.2    22.6   1836  6  ABL01352            Abl01352 Murine ap
c  12     28.6    22.2   1146  11  ABD03998            Abd03998 Pseudomon
   13     28.6    22.2   2184  11  ABD03637            Abd03637 Pseudomon
   14     28.6    22.2   2868  11  ABD03707            Abd03707 Pseudomon
   15     28.4    22.0    936  2  AAT79329            Aat79329 DNA encod
   16     28.4    22.0    936  8  ABX11507            Abx11507 DNA encod
   17     28.4    22.0    936  10  AAD58914            Aad58914 Archaeogl
   18     28.4    22.0   2451  6  ABZ12259            Abz12259 Arabidops
c  19     28.2    21.9    599  12  ACH79902            Ach79902 Human gen
   20     28.2    21.9   1152  5  AAS73582            Aas73582 DNA encod
```

```
   21     28.2    21.9  13540  8  ABZ74138            Abz74138 Secreted
   22     28.2    21.9  13540  8  ADA98673            Ada98673 Human sec
   23     28.2    21.9  13540  10  ABT16951            Abt16951 Human sec
   24     28.2    21.9  13540  10  ABZ67695            Abz67695 Human sec
c  25     28.2    21.9  18038  5  ADM20219            Adm20219 Alternati
   26     28.2    21.9  18038  8  ABZ74137            Abz74137 Secreted
   27     28.2    21.9  18038  8  ADA98672            Ada98672 Human sec
   28     28.2    21.9  18038  10  ABT16950            Abt16950 Human sec
   29     28.2    21.9  18038  10  ABZ67694            Abz67694 Human sec
c  30       28    21.7 165221  11  ACN44524            Acn44524 Mouse gen
c  31     27.8    21.6   2037  4  AAS46220            Aas46220 Human DNA
c  32     27.8    21.6   2037  8  ACA89670            Aca89670 cDNA enco
c  33     27.8    21.6   2037  8  ACA73680            Aca73680 Human sec
c  34     27.8    21.6   2037  8  ACA05995            Aca05995 Human sec
c  35     27.8    21.6   2037  8  ACA66829            Aca66829 cDNA enco
c  36     27.8    21.6   2037  8  ACF20404            Acf20404 Human sec
c  37     27.8    21.6   2037  8  ACF19790            Acf19790 Human sec
c  38     27.8    21.6   2037  8  ACD22078            Acd22078 Human sec
c  39     27.8    21.6   2037  8  ACF13243            Acf13243 Human sec
c  40     27.8    21.6   2037  8  ACD25346            Acd25346 Human sec
c  41     27.8    21.6   2037  8  ACF00395            Acf00395 Human sec
c  42     27.8    21.6   2037  8  ACA72452            Aca72452 Novel hum
c  43     27.8    21.6   2037  8  ACD04976            Acd04976 Novel hum
c  44     27.8    21.6   2037  8  ACD18437            Acd18437 Human sec
c  45     27.8    21.6   2037  8  ACD08444            Acd08444 Human sec

                               ALIGNMENTS

RESULT 1
ABK52324
ID   ABK52324 standard; DNA; 600 BP.
XX
AC   ABK52324;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding cancer gene therapy associated protein.
XX
KW   Solid cancer; anaerobic bacteria; antitumour active protein; cancer;
KW   gene therapy; gene; ds.
XX
OS   Bifidobacterium longum.
XX
FH   Key             Location/Qualifiers
FT   CDS             193. .474
FT                   /*tag=  a
FT                   /product= "Cancer gene therapy protein"
XX
PN   JP2002097144-A.
XX
PD   02-APR-2002.
XX
PF   21-SEP-2001; 2001JP-00290187.
XX
PR   21-SEP-2000; 2000JP-00287688.
XX
PA   (AMAN/) AMANO A.
PA   (FUJI/) FUJIMORI M.
XX
DR   WPI; 2002-448201/48.
DR   P-PSDB; AAU96807.
XX
PT   Solid cancer therapy with anaerobic bacteria of Bifidobacterium sp. by
PT   tumor tissue specific delivery of a DNA encoding for an antitumor active
PT   protein or its precursor.
XX
PS   Claim 10; Page 16; 21pp; Japanese.
XX
CC   The invention describes a method of treating a solid cancer with
CC   anaerobic bacteria by site specific delivery of DNA encoding an
CC   antitumour active protein or its precursor. This sequence encodes a
```

```
CC   cancer gene therapy associated protein
XX
SQ   Sequence 600 BP; 132 A; 165 C; 161 G; 142 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 129;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-34;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 2
ABQ81850/c
ID   ABQ81850 standard; DNA; 156638 BP.
XX
AC   ABQ81850;
XX
DT   19-NOV-2002  (first entry)
XX
DE   Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1106.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS   Bifidobacterium longum.
OS   Synthetic.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-00102050.
XX
PR   30-JAN-2001; 2001EP-00102050.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT   a probe or primer for detecting and/or identifying Bifidobacterium longum
PT   in a biological sample.
XX
PS   Disclosure; SEQ ID NO 1106; 80pp; English.
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC   fusion protein, comprising a sequence selected from 1097 sequences given
CC   in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC   heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC   activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC   a probe) is useful for the detection and/or identification of
CC   Bifidobacterium longum in a biological sample. A carrier containing the
CC   lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be
CC   used for preventing and/or treating diarrhoea brought about by pathogenic
CC   bacteria and/or rotavirus. The carrier is a food composition selected
CC   from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC   products, ice-creams, fermented cereal based products, milk based
```

```
CC   powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tablets, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification. N.B. The sequence data for this patent is not represented
CC   in the printed specification but is based on sequence information
CC   supplied by the European Patent Office
XX
SQ   Sequence 156638 BP; 32098 A; 46491 C; 46415 G; 31634 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 129;  DB 6;  Length 156638;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-33;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141691 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 141632

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141631 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 141572

Qy        121 AGTTCAGCG 129
              |||||||||
Db     141571 AGTTCAGCG 141563


RESULT 3
AAS59535
ID   AAS59535 standard; DNA; 26309 BP.
XX
AC   AAS59535;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein encoding DNA #30.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant; ds.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US012865.
XX
PR   21-APR-2000; 2000US-0199047P.
PR   02-JUN-2000; 2000US-0208841P.
PR   07-JUL-2000; 2000US-0216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris.
XX
PS   Claim 1; SEQ ID NO 30; 1069pp; English.
XX
CC   Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC   Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC   associated DNA sequences are used in the treatment, prevention and
```

```
CC   diagnosis of medical conditions caused by P. acnes. The disorders include
CC   SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC   osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
CC   infections of bone, joints and the central nervous system, however it is
CC   particularly involved in the inflammatory lesions associated with acne
CC   vulgaris. A method for detecting the presence or absence of P. acnes in a
CC   patient comprises contacting a sample with a binding agent that binds to
CC   the proteins of the invention and determining the amount of bound protein
CC   in the sample. The polypeptides may be used as antigens in the production
CC   of antibodies specific for P. acnes proteins. These antibodies can be
CC   used to downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC   polypeptides shown in AAU46704-AAU46985 and AAU67509. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;

  Query Match             23.7%;  Score 30.6;  DB 4;  Length 26309;
  Best Local Similarity   58.1%;  Pred. No. 9.8;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         37 TGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGT 96
              ||||||| |  | ||||||    ||  |  |||||  || | |||   |||||||   ||
Db       9538 TGACGACGACCATCCCGACTCCCATATTGAGGGTCGCTTCGATGTCAGGCTGTGAGACGT 9597

Qy         97 TGTCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              || ||| | |    |    |    || || |||
Db       9598 TGCCCACCTGCTGGACGAGGTTGAAGATCGGCG 9630


RESULT 4
ACF64464
ID   ACF64464 standard; DNA; 26309 BP.
XX
AC   ACF64464;
XX
DT   17-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes DNA contig sequence #30.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine; ds.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Claim 1; SEQ ID NO 30; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
```

```
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a P. acnes DNA contig which is specifically claimed
CC   in the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;

  Query Match             23.7%;  Score 30.6;  DB 8;  Length 26309;
  Best Local Similarity   58.1%;  Pred. No. 9.8;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         37 TGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGT 96
              ||||||| |  | ||||||    ||  |  |||||  || | |||   |||||||   ||
Db       9538 TGACGACGACCATCCCGACTCCCATATTGAGGGTCGCTTCGATGTCAGGCTGTGAGACGT 9597

Qy         97 TGTCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              || ||| | |    |    |    || || |||
Db       9598 TGCCCACCTGCTGGACGAGGTTGAAGATCGGCG 9630


RESULT 5
ADE63675
ID   ADE63675 standard; DNA; 2453 BP.
XX
AC   ADE63675;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human gene NM_004046, SEQ ID NO 9619.
XX
KW   Human; ds; gene; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; NM_004046.
```

```
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC   specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2453 BP; 622 A; 536 C; 635 G; 660 T; 0 U; 0 Other;

  Query Match             22.9%;  Score 29.6;  DB 10;  Length 2453;
  Best Local Similarity   57.6%;  Pred. No. 11;
  Matches   53;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || ||    || |  |||| ||| || ||  | || | ||||||   ||| ||   ||
Db       1864 CCATCACTGCGGACAGATTTCTTGGAACAGAATTGTCTACAAAGGATCCGCCCTCAATTA 1923

Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
               || ||||| |    | | || | |   ||||
Db       1924 ACGTGGTCTGTCGTATCTCGTTCGGATCCGTG 1955


RESULT 6
ADR91671
ID   ADR91671 standard; DNA; 885 BP.
XX
AC   ADR91671;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Novel S. pneumoniae DNA sequence, SEQ ID 306.
XX
KW   Meningitis; bacteraemia; pneumonia; otitis media; ds;
KW   bacterial infection.
XX
OS   Streptococcus pneumoniae.
XX
PN   US6800744-B1.
XX
PD   05-OCT-2004.
XX
PF   30-JUN-1998;   98US-00107433.
XX
PR   02-JUL-1997;   97US-0051553P.
PR   12-MAY-1998;   98US-0085131P.
```

```
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA,  Bush D;
XX
DR   WPI; 2004-697205/68.
DR   P-PSDB; ADR94274.
XX
PT   New isolated nucleic acid encoding a Streptococcus pneumoniae
PT   polypeptide, useful for diagnosing, preventing and/or treating
PT   pathological conditions resulting from the bacterial infection.
XX
PS   Disclosure; SEQ ID NO 306; 151pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a sequence
CC   encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC   fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC   ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC   ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC   ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC   ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC   sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC   hybridisable under high stringency conditions to the nucleotide sequence.
CC   The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC   Also included are a recombinant expression vector comprising the isolated
CC   nucleic acid cited above operably linked to a transcription regulatory
CC   element, a cell comprising the recombinant expression vector and a probe
CC   comprising at least 20 consecutive nucleotides of the nucleotide
CC   sequences as cited above. The methods and compositions of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   pathological conditions resulting from bacterial infection by
CC   Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC   otitis media. The present sequence is one of the 2603 disclosed S.
CC   pneumoniae nucleic acid sequences. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ   Sequence 885 BP; 243 A; 168 C; 218 G; 256 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 13;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 9.1;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        664 CGTGGTTTTGCCTGGTTGGATACGGGA 690


RESULT 7
ADQ89190/c
ID   ADQ89190 standard; DNA; 1377 BP.
XX
AC   ADQ89190;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Non-natural ND4 mitochondrial protein coding sequence.
XX
KW   gene therapy; ND4 mitochondrial protein; ND4; cellular dysfunction;
KW   mtDNA mutation; Leber Hereditary Optic Neuropathy;
KW   mitochondrial gene mutation; human; gene; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US2004142419-A1.
XX
PD   22-JUL-2004.
```

```
XX
PF   17-OCT-2003; 2003US-00687677.
XX
PR   18-OCT-2002; 2002US-0419435P.
XX
PA   (GUYJ/) GUY J.
XX
PI   Guy J;
XX
DR   WPI; 2004-579908/56.
XX
PT   New non-naturally occurring nucleic acid comprises a nucleotide sequence
PT   that encodes a functional ND4 mitochondrial protein, useful for reducing
PT   cellular dysfunction caused by mitochondrial gene mutations.
XX
PS   Claim 7; SEQ ID NO 1; 16pp; English.
XX
CC   The invention describes a non-naturally occurring nucleic acid comprising
CC   a nucleotide sequence that encodes a functional ND4 mitochondrial protein
CC   and differs from a naturally occurring nucleic acid that encodes a ND4
CC   mitochondrial protein by at least one codon substitution. Also described
CC   are: a cell into which has been introduced the non-naturally occurring
CC   nucleic acid above; and reducing dysfunction in a cell caused by a mtDNA
CC   mutation associated with Leber Hereditary Optic Neuropathy. Specifically
CC   claimed is non-naturally occurring ND4 nucleic acid comprising 1377 base
CC   pairs (SEQ ID NO. 1), fully defined in the specification. The nucleic
CC   acid is useful for reducing cellular dysfunction caused by mitochondrial
CC   gene mutations. Compositions comprising the non-naturally occurring
CC   nucleic acids are also useful for treating mtDNA mutations in animal
CC   subjects, including humans. This sequence represents a non-naturally
CC   occurring ND4 mitochondrial protein encoding polynucleotide sequence.
XX
SQ   Sequence 1377 BP; 330 A; 436 C; 341 G; 270 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 12;  Length 1377;
  Best Local Similarity   63.4%;  Pred. No. 10;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         57 GAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
              | || | | ||  || | ||||||  ||| |||| |  | || || || || || ||  |
Db       1292 GTGAAGCTGGGCTTCATGTTGTTGATGTGGTGTGTCAGGCTGCCCCACTGTGTTGTTGTG 1233

Qy        117 GACTAGTTCAG 127
               ||  || |||
Db       1232 AACATGTACAG 1222


RESULT 8
ADL45225
ID   ADL45225 standard; DNA; 3520 BP.
XX
AC   ADL45225;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human ovarian cancer DNA marker #19115.
XX
KW   Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS   Homo sapiens.
XX
PN   WO200170979-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US009126.
XX
PR   21-MAR-2000; 2000US-0191031P.
PR   25-MAY-2000; 2000US-0207124P.
PR   15-JUN-2000; 2000US-0211940P.
PR   07-JUL-2000; 2000US-0216820P.
PR   25-JUL-2000; 2000US-0220661P.
PR   21-DEC-2000; 2000US-0257672P.
```

```
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lee J,  Lillie J;
XX
DR   WPI; 2001-611502/70.
XX
PT   Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT   cancer cells as compared to their normal non-cancerous ovarian cells are
PT   used to characterize stage, grade, histological type of ovarian cancer.
XX
PS   Disclosure; SEQ ID NO 19115; 106pp; English.
XX
CC   The invention relates to nucleic acid markers which are overexpressed in
CC   ovarian cancer cells as compared to their expression in normal (i.e. non-
CC   cancerous) ovarian cells. The invention also relates to polypeptides
CC   encoded by the markers, antibodies that selectively bind to the
CC   polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC   of developing ovarian cancer involving inhibiting expression of a gene
CC   corresponding to a marker of the invention and a method of treating a
CC   patient afflicted with ovarian cancer comprising providing to cells of
CC   the patient an antisense oligonucleotide complementary to a marker of the
CC   invention. The markers are useful for assessing if a patient is afflicted
CC   with ovarian cancer, which involves comparing the level of expression of
CC   a marker in a patient sample and a normal level of expression of the
CC   marker in a control non-ovarian cancer sample. A difference between the
CC   expression levels indicates ovarian cancer. The level of expression of a
CC   marker corresponds to a secreted protein or to a transcribed
CC   polynucleotide or its portion. The level of expression of the marker is
CC   assessed by detecting the presence in the sample, a protein or protein
CC   fragment corresponding to the marker. The presence of protein or protein
CC   fragment is detected using an antibody that specifically binds with the
CC   protein or protein fragment. Alternatively, the level of expression of
CC   the marker is assessed by detecting the presence of a transcribed
CC   polynucleotide which anneals with the marker or anneals with a portion of
CC   the polynucleotide comprising the marker, under stringent conditions. The
CC   marker is also used for monitoring the progression of ovarian cancer in a
CC   patient which involves detecting expression of the marker in a patient
CC   sample at a first point in time, repeating the method at a subsequent
CC   time and comparing the level of expression. The method is carried out
CC   using an ovarian tissue sample. A composition comprising a marker,
CC   polypeptide or antibody of the invention is used to treat ovarian cancer.
CC   This sequence represents a human ovarian cancer DNA marker of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3520 BP; 679 A; 939 C; 1047 G; 855 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 5;  Length 3520;
  Best Local Similarity   54.1%;  Pred. No. 14;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || ||| ||  |      | |||      ||  | |    |   | |||   |  | |||
Db         27 CCCTCTACTGCTCATCTCTGCTTACCTTGTTTTTTAGTGGATTAATCCCTCACCTGGTGG 86

Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTC 114
               ||  ||  |||| |||||| ||||| || || || || ||   || || |
Db         87 GCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTCATAATCC 137


RESULT 9
ABL01350/c
ID   ABL01350 standard; DNA; 634 BP.
XX
AC   ABL01350;
XX
DT   15-MAR-2002  (first entry)
XX
DE   Murine apoptosis related DNA sequence #15.
XX
KW   Apoptosis; mouse; cancer; autoimmune disease; viral infection;
```

```
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   reperfusion injury; stroke; liver damage; dilatory cardiomyopathy;
KW   transgenic animal; hepatotropic; antialcoholism; cytostatic;
KW   immunosuppressive; virucide; nootropic; neuroprotective; vasotropic;
KW   antiparkinsonian; cerebroprotective; ds.
XX
OS   Mus sp.
XX
PN   DE10126344-A1.
XX
PD   24-JAN-2002.
XX
PF   30-MAY-2001; 2001DE-01026344.
XX
PR   14-JUL-2000; 2000DE-01034303.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Grimm S,  Schoenfeld N,  Braziulis E,  Cramer U,  Gewies A,  Voss F;
PI   Mund T,  Albayrak T,  Gille H,  Klein M;
XX
DR   WPI; 2002-115563/16.
XX
PT   New apoptosis-associated nucleic acid sequences and polypeptides, useful
PT   for diagnosis, treatment and prevention of e.g. tumors and
PT   neurodegeneration.
XX
PS   Claim 1; Page 137; 227pp; German.
XX
CC   The present invention relates to nucleic acids from the mouse, where the
CC   nucleic acid is associated with apoptosis. The sequences can be used in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   excessive or inadequate apoptosis, including tumours, autoimmune
CC   diseases, viral infections, degradative diseases (Alzheimer's,
CC   Parkinson's and Huntington's diseases), reperfusion injury, stroke and
CC   alcohol-induced injury to the liver, for identifying agents for treating
CC   these diseases, and to prepare transgenic animals in which expression of
CC   an apoptosis related sequence is altered. These are useful for genetic
CC   and/or pharmacological investigations of apoptosis and related diseases,
CC   including dilatory cardiomyopathy. The present sequence is one of the
CC   apoptosis related sequences of the invention
XX
SQ   Sequence 634 BP; 157 A; 184 C; 160 G; 133 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 6;  Length 634;
  Best Local Similarity   59.8%;  Pred. No. 9.7;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        245 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 186

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        185 ATCAATGCCTACTTGGTGGACG 164


RESULT 10
ADR89493
ID   ADR89493 standard; DNA; 1800 BP.
XX
AC   ADR89493;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Apoptosis-inducing protein coding sequence, SEQ ID 17.
XX
KW   Cytostatic; Immunosuppressive; Neuroprotective; Vasotropic; Virucide;
KW   apoptosis; neurodegenerative disease; ischaemic disease; cancer;
KW   autoimmune disease; viral disease; murine; gene; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1170
FT                   /*tag=  a
FT                   /product= "Apoptosis-inducing protein"
XX
PN   WO2004078112-A2.
XX
PD   16-SEP-2004.
XX
PF   05-MAR-2004; 2004WO-JP002899.
XX
PR   07-MAR-2003; 2003JP-00061179.
PR   10-MAR-2003; 2003US-0452943P.
XX
PA   (ASAH-) ASAHI KASEI PHARMA CORP.
XX
PI   Muramatsu S,  Takeda M,  Matsuda A;
XX
DR   WPI; 2004-662343/64.
DR   P-PSDB; ADR89494.
XX
PT   New protein capable of inducing apoptosis, useful in screening for
PT   compounds that inhibit or induce apoptosis which may be used to treat
PT   neurodegenerative, ischemic, autoimmune and viral diseases, and cancer.
XX
PS   Claim 4; SEQ ID NO 17; 316pp; English.
XX
CC   The present invention relates to novel purified proteins (I) and their
CC   coding sequences (II) (ADR89477-ADR89550), which are capable of inducing
CC   apoptosis. The proteins (I) are useful as a target in screening for
CC   compounds that modulate apoptosis. Compounds that modulate the expression
CC   or activity of the protein may be useful in treating neurodegenerative
CC   diseases, ischaemic diseases, cancer, autoimmune diseases, or viral
CC   diseases. Measurement of the expression or activity of the protein may
CC   also be used to diagnose or disease or a susceptibility to a disease.
XX
SQ   Sequence 1800 BP; 323 A; 526 C; 556 G; 395 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 13;  Length 1800;
  Best Local Similarity   59.8%;  Pred. No. 13;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        769 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 828

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        829 ATCAATGCCTACTTGGTGGACG 850


RESULT 11
ABL01352
ID   ABL01352 standard; DNA; 1836 BP.
XX
AC   ABL01352;
XX
DT   15-MAR-2002  (first entry)
XX
DE   Murine apoptosis related DNA sequence #17.
XX
KW   Apoptosis; mouse; cancer; autoimmune disease; viral infection;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   reperfusion injury; stroke; liver damage; dilatory cardiomyopathy;
KW   transgenic animal; hepatotropic; antialcoholism; cytostatic;
KW   immunosuppressive; virucide; nootropic; neuroprotective; vasotropic;
KW   antiparkinsonian; cerebroprotective; ds.
XX
OS   Mus sp.
XX
PN   DE10126344-A1.
XX
PD   24-JAN-2002.
```

```
XX
PF   30-MAY-2001; 2001DE-01026344.
XX
PR   14-JUL-2000; 2000DE-01034303.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Grimm S,  Schoenfeld N,  Braziulis E,  Cramer U,  Gewies A,  Voss F;
PI   Mund T,  Albayrak T,  Gille H,  Klein M;
XX
DR   WPI; 2002-115563/16.
XX
PT   New apoptosis-associated nucleic acid sequences and polypeptides, useful
PT   for diagnosis, treatment and prevention of e.g. tumors and
PT   neurodegeneration.
XX
PS   Claim 1; Page 138-139; 227pp; German.
XX
CC   The present invention relates to nucleic acids from the mouse, where the
CC   nucleic acid is associated with apoptosis. The sequences can be used in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   excessive or inadequate apoptosis, including tumours, autoimmune
CC   diseases, viral infections, degradative diseases (Alzheimer's,
CC   Parkinson's and Huntington's diseases), reperfusion injury, stroke and
CC   alcohol-induced injury to the liver, for identifying agents for treating
CC   these diseases, and to prepare transgenic animals in which expression of
CC   an apoptosis related sequence is altered. These are useful for genetic
CC   and/or pharmacological investigations of apoptosis and related diseases,
CC   including dilatory cardiomyopathy. The present sequence is one of the
CC   apoptosis related sequences of the invention
XX
SQ   Sequence 1836 BP; 341 A; 534 C; 562 G; 399 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 6;  Length 1836;
  Best Local Similarity   59.8%;  Pred. No. 13;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
               | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db         813 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 872

Qy         108 ATTATTCCGGACTAGTTCAGCG 129
               || | | |  ||| | |   ||
Db         873 ATCAATGCCTACTTGGTGGACG 894


RESULT 12
ABD03998/c
ID   ABD03998 standard; DNA; 1146 BP.
XX
AC   ABD03998;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #2602.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
```

```
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO70427.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 2602; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 1146 BP; 251 A; 343 C; 378 G; 174 T; 0 U; 0 Other;

  Query Match             22.2%;  Score 28.6;  DB 11;  Length 1146;
  Best Local Similarity   57.1%;  Pred. No. 19;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          39 ACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
               ||||  ||||||| | ||||  | |  | | ||||  || |||  |||   | | ||  |
Db         906 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 847

Qy          99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
               | || ||     |  |    | || || |||
Db         846 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 816


RESULT 13
ABD03637
ID   ABD03637 standard; DNA; 2184 BP.
XX
AC   ABD03637;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #2241.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO70066.
XX
```

```
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 2241; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 2184 BP; 342 A; 706 C; 726 G; 410 T; 0 U; 0 Other;

  Query Match             22.2%;  Score 28.6;  DB 11;  Length 2184;
  Best Local Similarity   57.1%;  Pred. No. 23;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |  | | ||||  || |||  |||   | | ||  |
Db       1538 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 1597

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     |  |    | || || |||
Db       1598 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 1628


RESULT 14
ABD03707
ID   ABD03707 standard; DNA; 2868 BP.
XX
AC   ABD03707;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Pseudomonas aeruginosa polynucleotide #2311.
XX
KW   Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW   antibacterial.
XX
OS   Pseudomonas aeruginosa.
XX
PN   US6551795-B1.
XX
PD   22-APR-2003.
XX
PF   18-FEB-1999;   99US-00252991.
XX
PR   18-FEB-1998;   98US-0074788P.
PR   27-JUL-1998;   98US-0094190P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Rubenfield MJ,  Nolling J,  Deloughery C,  Bush D;
XX
DR   WPI; 2003-615309/58.
DR   P-PSDB; ABO70136.
XX
PT   Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT   useful as molecular targets for diagnostics, prophylaxis and treatment of
PT   pathological conditions resulting from bacterial infection.
XX
PS   Disclosure; SEQ ID NO 2311; 455pp; English.
XX
CC   The invention relates to Pseudomonas aeruginosa polypeptides and the
CC   polynucleotides encoding them. The sequences are useful in diagnosis and
CC   therapy of pathological conditions, as molecular targets for diagnostics,
CC   prophylaxis and treatment of pathological conditions resulting from a
CC   bacterial infection, for evaluating a compound, such as a polypeptide,
CC   for the ability to bind a P. aeruginosa nucleic acid, as components of
CC   effective antibacterial targets, as targets for antibacterial drugs,
CC   including anti-P. aeruginosa drugs, as templates for recombinant
CC   production of P. aeruginosa-derived peptides or polypeptides, as target
CC   components for diagnosis and/or treatment of P. aeruginosa-caused
CC   infection, and in detection of P. aeruginosa sequences or other sequences
CC   of Pseudomonas species using biochip technology. Sequences ABD01397-
CC   ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format from USPTO at
CC   seqdata.uspto.gov/sequence.html
XX
SQ   Sequence 2868 BP; 435 A; 939 C; 941 G; 553 T; 0 U; 0 Other;

  Query Match             22.2%;  Score 28.6;  DB 11;  Length 2868;
  Best Local Similarity   57.1%;  Pred. No. 24;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |  | | ||||  || |||  |||   | | ||  |
Db       1177 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 1236

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     |  |    | || || |||
Db       1237 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 1267


RESULT 15
AAT79329
ID   AAT79329 standard; DNA; 936 BP.
XX
AC   AAT79329;
XX
DT   17-OCT-2003  (revised)
DT   16-FEB-1998  (first entry)
XX
DE   DNA encoding Archaeoglobus fulgidus esterase VC16-16MC.
XX
KW   Esterase; thermostable enzyme; ester; chiral compound; cheese; pulp;
KW   paper; lignin removal; sugar; lignocellulose; disease resistance;
KW   feedstuff; ss.
XX
OS   Archaeoglobus fulgidus; strain VC16.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .936
FT                   /*tag=  b
FT                   /transl_except= (pos: 382. .384, aa:Ala)
FT                   /transl_except= (pos: 385. .387, aa:Ala)
FT                   /transl_except= (pos: 430. .432, aa:Ala)
FT                   /transl_except= (pos: 487. .489, aa:Gly)
FT                   /transl_except= (pos: 499. .501, aa:Ala)
FT                   /transl_except= (pos: 505. .507, aa:Val)
FT                   /transl_except= (pos: 559. .561, aa:Ile)
FT                   /transl_except= (pos: 691. .693, aa:Asn)
FT   unsure          313. .315
FT                   /*tag=  a
FT                   /note= "encodes Ile"
XX
PN   WO9730160-A1.
XX
PD   21-AUG-1997.
XX
PF   11-FEB-1997;   97WO-US002039.
XX
PR   16-FEB-1996;   96US-00602359.
```

```
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Robertson DE,  Murphy D,  Reid J,  Maffia AM,  Link S,  Swanson RV;
PI   Warren PV,  Kosmotka A,  Callen W;
XX
DR   WPI; 1997-425035/39.
DR   P-PSDB; AAW23076.
XX
PT   Nucleic acid encoding heat stable esterase from thermophilic bacteria -
PT   which is active in organic solvents, useful in cheese or paper
PT   manufacture, and to study plant resistance to disease.
XX
PS   Claim 1; Page 50-51; 113pp; English.
XX
CC   This DNA sequence codes for thermostable esterase VC16-16MC (AAW23076) of
CC   Archaeoglobus fulgidus VC16, an isolate that grows optimally at 85 deg C
CC   and pH 7.0. It can be amplified from a pBluescript vector by PCR (see
CC   AAT79317-18). Claimed, newly identified polynucleotides (AAT79321-30)
CC   encoding esterases (AAW23069-77, AAW23088) were recovered from genomic
CC   gene libraries. They can be used for recombinant production of the
CC   enzymes in host cells, and as probes to identify related sequences. The
CC   esterases are stable at high temperature and in organic solvents, making
CC   them superior for use in production of pure chiral compounds used in
CC   pharmaceutical, agricultural and other chemical industries. A method is
CC   claimed for transferring an amino group from an amino acid to an alpha-
CC   keto acid using a claimed esterase. The enzymes may also be useful as
CC   ripening starters in cheese making, in lignin removal in paper and pulp
CC   manufacture, in carbohydrate derivative synthesis, in fermentable sugar
CC   production from lignocellulosic waste, in the study of plant wall
CC   structure, plant resistance to disease and organic matter decomposition
CC   and to select plants bred for production of highly degradable animal
CC   feeds. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 936 BP; 235 A; 220 C; 265 G; 214 T; 0 U; 2 Other;

  Query Match             22.0%;  Score 28.4;  DB 2;  Length 936;
  Best Local Similarity   56.4%;  Pred. No. 20;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              ||||||  |  | || |   | |   || ||| || | |||  | |||||     |||||
Db        718 GACCTTGAGAACCTACCTCCTGCGCTGATCATAACCGCCGAATACGACCCGCTGAGAGAT 777

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
              |   | |   |||| |    | ||||| || | |
Db        778 GAAGGAGAAGTTTTCGGGCAGATGCTGAGAAGAG 811


Search completed: August  5, 2005, 23:52:43
Job time : 209.159 secs
```

```
                        GenCore version 5.1.6
               Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          August  5, 2005, 23:28:19 ; Search time 68.7196 Seconds
                                              (without alignments)
                                              3071.607 Million cell updates/sec

Title:           US-10-782-899-1_COPY_472_600
Perfect score:   129
Sequence:        1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:      2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                        %
Result                Query
   No.      Score     Match Length DB  ID                    Description
-------------------------------------------------------------------------
     1       29.4      22.8     870  4  US-09-583-110-1324    Sequence 1324, Ap
     2       29.4      22.8     885  4  US-09-107-433-306     Sequence 306, App
     3       29.4      22.8  194915  4  US-09-949-016-15584   Sequence 15584, A
     4         29      22.5     601  4  US-09-949-016-68059   Sequence 68059, A
     5         29      22.5  264665  4  US-09-949-016-13747   Sequence 13747, A
c    6       28.6      22.2    1146  4  US-09-252-991A-2602   Sequence 2602, Ap
     7       28.6      22.2    2184  4  US-09-252-991A-2241   Sequence 2241, Ap
     8       28.6      22.2    2868  4  US-09-252-991A-2311   Sequence 2311, Ap
     9       28.4      22.0     936  2  US-08-602-359A-31     Sequence 31, Appl
c   10       27.4      21.2    1160  4  US-09-270-767-13705   Sequence 13705, A
c   11       26.6      20.6     601  4  US-09-949-016-125432  Sequence 125432,
    12       26.6      20.6  139552  4  US-09-949-016-15300   Sequence 15300, A
c   13       26.4      20.5   14707  4  US-09-312-762A-3      Sequence 3, Appli
c   14       26.2      20.3     601  4  US-09-949-016-143314  Sequence 143314,
c   15       26.2      20.3     601  4  US-09-949-016-143485  Sequence 143485,
c   16       26.2      20.3     746  4  US-09-380-419C-1      Sequence 1, Appli
c   17       26.2      20.3     907  3  US-08-672-850-7       Sequence 7, Appli
c   18       26.2      20.3     907  3  US-08-672-850-11      Sequence 11, Appl
c   19       26.2      20.3     907  4  US-09-565-177A-7      Sequence 7, Appli
c   20       26.2      20.3     907  4  US-09-565-177A-11     Sequence 11, Appl
c   21       26.2      20.3    1001  3  US-08-672-850-10      Sequence 10, Appl
c   22       26.2      20.3    1001  4  US-09-565-177A-10     Sequence 10, Appl
    23       26.2      20.3   85912  4  US-09-949-016-12362   Sequence 12362, A
    24       26.2      20.3   85913  4  US-09-949-016-16109   Sequence 16109, A
    25       26.2      20.3  120727  4  US-09-949-016-15787   Sequence 15787, A
    26       26.2      20.3  120727  4  US-09-949-016-15788   Sequence 15788, A
c   27         26      20.2    1074  4  US-09-934-901-7       Sequence 7, Appli
```

```
c   28         26      20.2    1074  4  US-09-934-868-17      Sequence 17, Appl
c   29         26      20.2    1074  4  US-10-321-210-7       Sequence 7, Appli
c   30         26      20.2    1074  4  US-10-320-874-7       Sequence 7, Appli
c   31         26      20.2   43280  2  US-08-804-227C-1      Sequence 1, Appli
    32         26      20.2  336024  4  US-09-949-016-12373   Sequence 12373, A
c   33       25.8      20.0    1158  3  US-09-198-092-1       Sequence 1, Appli
c   34       25.8      20.0    2262  2  US-08-674-887A-5      Sequence 5, Appli
c   35       25.8      20.0    2262  3  US-08-951-844-5       Sequence 5, Appli
c   36       25.8      20.0    2262  3  US-09-412-347-5       Sequence 5, Appli
    37       25.6      19.8     271  4  US-09-313-294A-1215   Sequence 1215, Ap
    38       25.6      19.8   46725  4  US-09-949-016-15680   Sequence 15680, A
    39       25.6      19.8  183112  4  US-09-949-016-14184   Sequence 14184, A
c   40       25.6      19.8  481115  4  US-09-949-016-11940   Sequence 11940, A
c   41       25.4      19.7    1158  4  US-09-023-655-992     Sequence 992, App
c   42       25.4      19.7    1586  1  US-08-461-244-1       Sequence 1, Appli
c   43       25.4      19.7    1953  4  US-09-016-434-1096    Sequence 1096, Ap
c   44       25.4      19.7    2608  4  US-09-023-655-955     Sequence 955, App
    45       25.4      19.7    3302  3  US-09-221-017B-174    Sequence 174, App


                                  ALIGNMENTS


RESULT 1
US-09-583-110-1324
; Sequence 1324, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 1324
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-1324

  Query Match             22.8%;  Score 29.4;  DB 4;  Length 870;
  Best Local Similarity   58.6%;  Pred. No. 0.94;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        589 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 648

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        649 CGTGGTTTTGCCTGGTTGGATACGGGA 675


RESULT 2
US-09-107-433-306
; Sequence 306, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;         APPLICANT: Lynn A Doucette-Stamm  and David Bush
;         TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                             SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGN
;                             THERAPEUTICS
;         NUMBER OF SEQUENCES: 5206
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;              STREET: 100 Beaver Street
```

```
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: <Unknown>
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: <Unknown>
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,433
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/ 085131
;            FILING DATE: May 12, 1998
;            APPLICATION NUMBER: 60/051553
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 306:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 885 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Streptococcus pneumoniae
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...885
;       SEQUENCE DESCRIPTION: SEQ ID NO: 306:
US-09-107-433-306

  Query Match             22.8%;  Score 29.4;  DB 4;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 0.95;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        664 CGTGGTTTTGCCTGGTTGGATACGGGA 690


RESULT 3
US-09-949-016-15584
; Sequence 15584, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 15584
;   LENGTH: 194915
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(194915)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15584

  Query Match             22.8%;  Score 29.4;  DB 4;  Length 194915;
  Best Local Similarity   60.8%;  Pred. No. 10;
  Matches   48;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy         42 ACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCC 101
              | |||  | ||    ||||     | |  || |||||||||||||| ||  || ||| |
Db     115691 AGAAACTCACCACTTGAGACAAAAGTGAACTCTTTGTTGTGGTGCTATGTGGTCTTGGCA 115750

Qy        102 AACCGTATTATTCCGGACT 120
               |   |  | |||||| ||
Db     115751 GAGGTTTGTTTTCCGGTCT 115769


RESULT 4
US-09-949-016-68059
; Sequence 68059, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68059
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-68059

  Query Match             22.5%;  Score 29;  DB 4;  Length 601;
  Best Local Similarity   58.8%;  Pred. No. 1.1;
  Matches   50;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         23 ACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTG 82
              | | | || ||||| | ||  |   |||  | |||    |  | || ||||| | ||||
Db         30 AGTGCTAGGATTACAGGCGTGAGCCACCGTGCCCGGCCAGACCTGGCTCTTTCTTTTGTT 89

Qy         83 GTGCTGTGACGTGTTGTCCAACCGT 107
               |  || ||||   |||||  | ||
Db         90 TTTTTGAGACGGAGTGTCCCTCTGT 114


RESULT 5
US-09-949-016-13747
; Sequence 13747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
```

```
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13747
;   LENGTH: 264665
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13747

  Query Match             22.5%;  Score 29;  DB 4;  Length 264665;
  Best Local Similarity   58.8%;  Pred. No. 17;
  Matches   50;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         23 ACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGTCTTTTTGTTGTG 82
              | | | || ||||| | ||  |    |||  | |||      |   | || ||||| | ||||
Db      52981 AGTGCTAGGATTACAGGCGTGAGCCACCGTGCCCGGCCAGACCTGGCTCTTTCTTTTGTT 53040

Qy         83 GTGCTGTGACGTGTTGTCCAACCGT 107
               |  || ||||   |||||  | ||
Db      53041 TTTTTGAGACGGAGTGTCCCTCTGT 53065


RESULT 6
US-09-252-991A-2602/c
; Sequence 2602, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2602
;   LENGTH: 1146
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2602

  Query Match             22.2%;  Score 28.6;  DB 4;  Length 1146;
  Best Local Similarity   57.1%;  Pred. No. 2.1;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |   | | ||||  || |||  |||   | | || |
Db        906 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 847

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     | |    | || || |||
Db        846 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 816


RESULT 7
US-09-252-991A-2241
; Sequence 2241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2241
;   LENGTH: 2184
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2241

  Query Match             22.2%;  Score 28.6;  DB 4;  Length 2184;
  Best Local Similarity   57.1%;  Pred. No. 2.8;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |   | | ||||  || |||  |||   | | || |
Db       1538 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 1597

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     | |    | || || |||
Db       1598 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 1628


RESULT 8
US-09-252-991A-2311
; Sequence 2311, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2311
;   LENGTH: 2868
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2311

  Query Match             22.2%;  Score 28.6;  DB 4;  Length 2868;
  Best Local Similarity   57.1%;  Pred. No. 3.2;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |   | | ||||  || |||  |||   | | || |
Db       1177 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 1236

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     | |    | || || |||
Db       1237 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 1267


RESULT 9
US-08-602-359A-31
; Sequence 31, Application US/08602359A
; Patent No. 5942430
;  GENERAL INFORMATION:
;     APPLICANT:  ROBERTSON, Daniel E.
;     APPLICANT:  MURPHY, Dennis
;     APPLICANT:  REID, John
;     APPLICANT:  MAFFIA, Anthony
;     APPLICANT:  LINK, Steven
;     APPLICANT:  SWANSON, Ronald V.
;     APPLICANT:  WARREN, Patrick V.
;     APPLICANT:  KOSMOTKA, Anna
;     TITLE OF INVENTION:  ESTERASES
;     NUMBER OF SEQUENCES:  42
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 EXECUTIVE SQUARE, STE 1400
;      CITY:  LA JOLLA
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,359A
;      FILING DATE:  February 16, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-678-5070
;      TELEFAX:  619-678-5099
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  GENOMIC DNA
US-08-602-359A-31

  Query Match             22.0%;  Score 28.4;  DB 2;  Length 936;
  Best Local Similarity   56.4%;  Pred. No. 2.3;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              ||||||  |  | || |   | |   || ||| || | |||  | |||||     |||||
Db        718 GACCTTGAGAACCTACCTCCTGCGCTGATCATAACCGCCGAATACGACCCGCTGAGAGAT 777

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
              |   | |   |||| |    | ||||| || | |
Db        778 GAAGGAGAAGTTTTCGGGCAGATGCTGAGAAGAG 811


RESULT 10
US-09-270-767-13705/c
; Sequence 13705, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13705
;   LENGTH: 1160
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-13705

  Query Match             21.2%;  Score 27.4;  DB 4;  Length 1160;
  Best Local Similarity   54.5%;  Pred. No. 5.9;
  Matches   55;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || | | |  || |    || ||| | |     ||   | ||| ||| ||| |  ||||
Db        656 CCATTTTCGAGTCGTCGCTAATTCCATCCACCTTGTGTATAAACACCTCGAACTTGATGC 597
```

```
Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAAC 104
               |  | ||  ||||| |   ||| |    ||||| |  |||
Db        596 GCTTGTTCACTTTGTAGGCCTGCAGCACTGTGTTCTTGAAC 556


RESULT 11
US-09-949-016-125432/c
; Sequence 125432, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 125432
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-125432

  Query Match             20.6%;  Score 26.6;  DB 4;  Length 601;
  Best Local Similarity   66.7%;  Pred. No. 8.6;
  Matches   38;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         44 AAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              |||| | | ||  ||||||     |||||| |   | ||| ||||| | || |||||
Db        407 AAAGCCTCTGATTGAGATGAGGAAGGTCTTCTACCTCTGGAGCTGTAAGGTCTTGTC 351


RESULT 12
US-09-949-016-15300
; Sequence 15300, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15300
;   LENGTH: 139552
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15300

  Query Match             20.6%;  Score 26.6;  DB 4;  Length 139552;
  Best Local Similarity   66.7%;  Pred. No. 97;
  Matches   38;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         44 AAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              |||| | | ||  ||||||     |||||| |   | ||| ||||| | || |||||
Db      80719 AAAGCCTCTGATTGAGATGAGGAAGGTCTTCTACCTCTGGAGCTGTAAGGTCTTGTC 80775
```

```
RESULT 13
US-09-312-762A-3/c
; Sequence 3, Application US/09312762A
; Patent No. 6552177
;  GENERAL INFORMATION:
;    APPLICANT:  MIA HOROWITZ ET AL.
;    TITLE OF INVENTION:  EH DOMAIN CONTAINING GENES AND PROTEINS
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mark M. Friedman c/o Anthony Castorina
;      STREET:  2001 Jefferson Davis Highway, Suite 207
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  United States of America
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  1.44 megabyte, 3.5" microdisk
;      COMPUTER:  Twinhead* Slimnote-890TX
;      OPERATING SYSTEM:  MS DOS version 6.2,
;      OPERATING SYSTEM:  Windows version 3.11
;      SOFTWARE:  Word for Windows version 2.0 converted to
;      SOFTWARE:  an ASCI file
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/312,762A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/026,898
;      FILING DATE:  20 FEB 1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Friedmam, Mark M.
;      REGISTRATION NUMBER:  33,883
;      REFERENCE/DOCKET NUMBER:  916/10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  972-3-5625553
;      TELEFAX:  972-3-5625554
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14707
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-312-762A-3

  Query Match             20.5%;  Score 26.4;  DB 4;  Length 14707;
  Best Local Similarity   54.0%;  Pred. No. 42;
  Matches   54;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          7 TCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCG 66
              |||  || | | ||||  |  | || ||  ||||  |  ||| | || | ||||||  |
Db       5137 TCTTGTCCTCGTGATTTTTGAGGGCCTTGATGACTTCTGAGAACTCGTCTGAGATGTCCA 5078

Qy         67 GGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCG 106
              |  | |    || |     |   ||  ||  |||||  ||
Db       5077 GCTTGTGGGCGTCGAACAACAAAGAGATGCGGTCCACACG 5038

RESULT 14
US-09-949-016-143314/c
; Sequence 143314, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143314
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-143314

  Query Match             20.3%;  Score 26.2;  DB 4;  Length 601;
  Best Local Similarity   58.2%;  Pred. No. 12;
  Matches   46;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         29 AGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTG 88
              ||| |  |||||  | | | || | || |   |||| |||  |  | |     | | |||
Db        291 AGCTTCTCTGACACCCACGCCCACCACAGTCTTGGTTGGGTGCCCTGTCCCCAGATCCTG 232

Qy         89 TGACGTGTTGTCCAACCGT 107
              || | | |||| || | ||
Db        231 TGGCTTTTTGTTCACCTGT 213

RESULT 15
US-09-949-016-143485/c
; Sequence 143485, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143485
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-143485

  Query Match             20.3%;  Score 26.2;  DB 4;  Length 601;
  Best Local Similarity   58.2%;  Pred. No. 12;
  Matches   46;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         29 AGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTG 88
              ||| |  |||||  | | | || | || |   |||| |||  |  | |     | | |||
Db        291 AGCTTCTCTGACACCCACGCCCACCACAGTCTTGGTTGGGTGCCCTGTCCCCAGATCCTG 232

Qy         89 TGACGTGTTGTCCAACCGT 107
              || | | |||| || | ||
Db        231 TGGCTTTTTGTTCACCTGT 213


Search completed: August  6, 2005, 01:33:33
Job time : 70.7196 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 22:54:59 ; Search time 1309.69 Seconds
                                            (without alignments)
                                            3749.196 Million cell updates/sec

Title:          US-10-782-899-1_COPY_472_600
Perfect score:  129
Sequence:       1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:      68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                   %
Result             Query
  No.   Score      Match Length DB  ID                  Description
-------------------------------------------------------------------------------
c   1    31.8      24.7    600  4   BG817831            BG817831 ESSU0253
    2      31      24.0    460  7   CF927863            CF927863 laf58f02.
c   3      31      24.0    932  7   CK279022            CK279022 EST725100
c   4    30.8      23.9    429  4   BM522110            BM522110 ESSU0647
    5    30.8      23.9    524  1   AA517086            AA517086 vh98f10.r
c   6    30.8      23.9    938  7   CN808355            CN808355 Blood_EST
c   7    30.6      23.7    199  1   AJ282728            AJ282728 4A3A-ABB-
    8    30.6      23.7    677  1   AV745983            AV745983 AV745983
c   9    30.2      23.4    494  4   BM522137            BM522137 ESSU0674
c  10    30.2      23.4    528  4   BM522043            BM522043 ESSU0580
c  11    30.2      23.4    539  4   BM522233            BM522233 ESSU0771
c  12    30.2      23.4    546  4   BG817800            BG817800 ESSU0222
c  13    30.2      23.4    572  4   BM522287            BM522287 ESSU0825
c  14    30.2      23.4    577  4   BG817916            BG817916 ESSU0338
c  15    30.2      23.4    599  4   BG817905            BG817905 ESSU0327
c  16    30.2      23.4    657  4   BG817854            BG817854 ESSU0276
c  17    30.2      23.4    682  4   BG817862            BG817862 ESSU0284
c  18    30.2      23.4    720  4   BG817630            BG817630 ESSU0052
c  19    30.2      23.4   1101  9   CNS017UC            AL108510 Drosophil
c  20    29.8      23.1    125  6   CA829999            CA829999 3529_1_3_
c  21    29.8      23.1    486  6   CB816450            CB816450 3529_1_79
   22    29.8      23.1    655  7   CK561781            CK561781 rswpb0_00
   23    29.8      23.1   1396  6   CD326834            CD326834 AGENCOURT
c  24    29.6      22.9    187  2   BF841751            BF841751 RC2-HT107
c  25    29.6      22.9    403  4   BG925693            BG925693 HNC13-1-B
c  26    29.6      22.9    819  7   CN168123            CN168123 AGENCOURT
   27    29.6      22.9    903  9   CNS04GSE            AL290039 Tetraodon
c  28    29.6      22.9    937  6   CD107403            CD107403 AGENCOURT
c  29    29.4      22.8    137  4   BJ305960            BJ305960 BJ305960
c  30    29.4      22.8    437  4   BM521973            BM521973 ESSU0510
   31    29.4      22.8    485  7   H76502              H76502 18207 Lambd
   32    29.4      22.8    770  9   CG997493            CG997493 ZMMBBb053
   33    29.4      22.8   1353  4   BI114606            BI114606 602861984
   34    29.2      22.6    259  1   AI613789            AI613789 vh98f10.y
   35    29.2      22.6    345  2   BE945901            BE945901 UI-M-BZ0-
   36    29.2      22.6    356  6   CB813539            CB813539 AMGNNUC:M
   37    29.2      22.6    365  7   W48020              W48020 mc91e03.r1
   38    29.2      22.6    406  2   BE651993            BE651993 UI-M-AP0-
   39    29.2      22.6    406  6   CB807606            CB807606 AMGNNUC:S
   40    29.2      22.6    406  7   CO059265            CO059265 est_k_bre
   41    29.2      22.6    411  6   CB802272            CB802272 AMGNNUC:N
   42    29.2      22.6    415  1   AA103400            AA103400 mo24f02.r
   43    29.2      22.6    420  2   BE226847            BE226847 us73d03.y
   44    29.2      22.6    476  6   CB732090            CB732090 AMGNNUC:N
   45    29.2      22.6    518  7   CO429424            CO429424 UI-M-HW0-
```

ALIGNMENTS

```
RESULT 1
BG817831/c
LOCUS       BG817831                 600 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  ESSU0253 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0334 5', mRNA sequence.
ACCESSION   BG817831
VERSION     BG817831.1  GI:14188811
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to pir|T17405 scavenger recept or cysteine-rich protein
            precursor - sea urchin (Strongyloce ntrotus purpuratus).
            Seq primer: T3 primer
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0334"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             24.7%;  Score 31.8;  DB 4;  Length 600;
  Best Local Similarity   57.6%;  Pred. No. 13;
  Matches   57;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;
```

```
Qy            2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
                |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db          423 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 364

Qy           62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
                | |  |||  | | || || |||  | ||   |||| ||
Db          363 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTTTTC 325


RESULT 2
CF927863
LOCUS       CF927863                 460 bp    mRNA    linear   EST 12-NOV-2003
DEFINITION  laf58f02.y1 SiEP Mus musculus cDNA 5' similar to TR:Q9UJZ5 Q9UJZ5
            PRESENILIN-ASSOCIATED PROTEIN. ;, mRNA sequence.
ACCESSION   CF927863
VERSION     CF927863.1  GI:38274486
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Tidwell,R., Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J.,
            Wylie,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennet,J.,
            Ronko,I., Tsagareishvili,R., Belaygorod,L., Grow,A., Maguire,L.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Stem cell EST Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Jeff Gordon and Mike Lovett
            WashU, Human Genetics Division
            Washington University School of Medicine
            Library materials provided by: Stappenbeck Th.S. and Gordon J.I.
            Library constructed by: Korshunova Y. and Lovett M. DNA sequencing
            by: Washington University Genome Sequencing Center For information
            on obtaining a clone please contact: Rose Tidwell
            (rtidwell@genetics.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1..460
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="small intestinal epithelial progenitors"
                     /lab_host="DH5alpha Ultra Max cells (Invitrogen)"
                     /clone_lib="SiEP"
                     /note="Vector: pAMP1; The library was synthesized with
                     modified SMART primers with dUTP at the end. After
                     treatment with UDG the cDNA was cloned in pAMP1 vector by
                     annealing."
ORIGIN

  Query Match             24.0%;  Score 31;  DB 7;  Length 460;
  Best Local Similarity   59.8%;  Pred. No. 24;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy           43 CAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCA 102
                | |  | |||   | | | || ||  ||  |||| |||||| ||||| || || || ||
Db            7 CTATTAACCCTCACTAAAGGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCC 66

Qy          103 ACCGTATTATTCCGGACTAGTTCAGCG 129
                ||   || | | |  ||| | |   ||
Db           67 ACTTCATCAATGCCTACTTGGTGGACG 93


RESULT 3
CK279022/c
LOCUS       CK279022                 932 bp    mRNA    linear   EST 03-AUG-2004
DEFINITION  EST725100 potato abiotic stress cDNA library Solanum tuberosum cDNA
            clone POAE963 5' end, mRNA sequence.
ACCESSION   CK279022
VERSION     CK279022.1  GI:39836000
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
  TITLE     Generation of ESTs from abiotic stressed potato tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from the University of Arizona Genomics
            Institute via http://genome.arizona.edu/orders/ .
            Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES             Location/Qualifiers
     source          1..932
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="POAE963"
                     /tissue_type="abiotic stress treated leaf and root tissue"
                     /lab_host="DH10B-TonA"
                     /clone_lib="potato abiotic stress cDNA library"
                     /note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI;
                     supplier: Solanum tuberosum var. Kennebec plants were
                     grown from cuttings on a 16hr light/8 hr dark cycle at 25
                     C for 3-4 weeks. Abiotic stress conditions were applied to
                     four separate sets of plants. Set 1 involved saturation of
                     the soil with 150 mM NaCl and tissues were harvested at
                     following application of the salt stress (leaves: 2hr,
                     6hr, 12hr, 1d, 2d, and 4d; roots:2hr, 6hr, 12hr, and 2d).
                     Set 2 were grown under the standard conditions and then
                     were water stressed by withdrawal of further watering
                     applications. Drought stressed plants were harvested after
                     cessation of watering (leaves: 3d, 5d, and 7d; roots:3d
                     and 5d). Set 3 were grown under the standard conditions
                     and then were cold stressed by placement at 4 C. Cold
                     stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
                     and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d,
                     2d. Set 4 were grown under the standard conditions and
                     then were heat stressed by placement at 35 C. Heat
                     stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
                     2d and 4d and heat-stressed roots were harvested at 6 hr,
                     12 hr, 1 d, and 4d. RNA was isolated from all tissues and
                     equal RNA from each tissue and stress was pooled to
                     construct the cDNA library. RNA sample."
ORIGIN

  Query Match             24.0%;  Score 31;  DB 7;  Length 932;
  Best Local Similarity   57.9%;  Pred. No. 27;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy            5 CTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGT 64
                |||||| |  ||  || ||    |   || | ||| || |||||    ||  | |||| |
Db          844 CTTCTGTTAATATTGAATAGAAAGCTGATGATTGAGGAGAAAGAGAGTGATGGTGATGAT 785

Qy           65 CGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGT 99
                   | | || ||||||| ||  |      ||||||
Db          784 GTTGTTGTTGTTGTTGTTGTAATTGATGTTGTTGT 750


RESULT 4
BM522110/c
LOCUS       BM522110                 429 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0647 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0777 5', mRNA sequence.
ACCESSION   BM522110
VERSION     BM522110.1  GI:18706032
KEYWORDS    EST.
```

```
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7495754|pir|T19130 hypothetical protein C09F9.2
            [Caenorhabditis elegans]
            Seq primer: T3 primer
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1..429
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0777"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.9%;  Score 30.8;  DB 4;  Length 429;
  Best Local Similarity   57.1%;  Pred. No. 27;
  Matches   56;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        406 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 347

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGT 99
              | |  |||  | | || || |||  | ||   |||| |
Db        346 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTTTT 309


RESULT 5
AA517086
LOCUS       AA517086                 524 bp    mRNA    linear   EST 14-JUL-1997
DEFINITION  vh98f10.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:902347 5', mRNA sequence.
ACCESSION   AA517086
VERSION     AA517086.1  GI:2256675
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:523011
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1..524
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:902347"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC,  catalog # CRL-1772) differentiates
                     rapidly, forming contractile  myotubes and producing
                     characteristic muscle proteins.   "
ORIGIN

  Query Match             23.9%;  Score 30.8;  DB 1;  Length 524;
  Best Local Similarity   61.0%;  Pred. No. 28;
  Matches   50;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||| |||| |||||| ||||| || || || || ||
Db         36 ATCCCTCACCTCCTGGGCGATGTCGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 95

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db         96 ATCAATGCCTACTTGCTGCACG 117


RESULT 6
CN808355/c
LOCUS       CN808355                 938 bp    mRNA    linear   EST 27-MAY-2004
DEFINITION  Blood_EST0439 Metarhizium anisopliae ARSEF 2575 from insect blood
            Metarhizium anisopliae cDNA clone B819 5', mRNA sequence.
ACCESSION   CN808355
VERSION     CN808355.1  GI:47729828
KEYWORDS    EST.
SOURCE      Metarhizium anisopliae
  ORGANISM  Metarhizium anisopliae
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
            Clavicipitaceae; Metarhizium.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   Wang,C.S., Hu,G. and St. Leger,R.J.
  TITLE     Gene expression profiling of Metarhizium anisopliae grown under
            different conditions: mechanisms of fungal opportunism
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Wang CS
            Department of Entomology
            University of Maryland
            4112 Plant Sciences Building, College Park, MD 20742, USA
            Email: cwang4@umd.edu
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..938
                     /organism="Metarhizium anisopliae"
                     /mol_type="mRNA"
                     /db_xref="taxon:5530"
```

```
                     /clone="B819"
                     /clone_lib="Metarhizium anisopliae ARSEF 2575 from insect
                     blood"
                     /note="Vector: pCMV.SPORT6.1; Metarhizium anisopliae was
                     grown in insect haemolymph for 24 hours. A cDNA library
                     was constructed in the vector pCMV.SPORT6.1"
ORIGIN

  Query Match             23.9%;  Score 30.8;  DB 7;  Length 938;
  Best Local Similarity   63.5%;  Pred. No. 31;
  Matches   47;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy           7 TCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCG 66
               ||||  ||  | |  | ||||  || || |||||||||||| |  || || |   |||   
Db         330 TCTGGGCGGCGGGGGTGCTTCCGGCGTTCCTGACGACAAAGTCGACGGCCGTCTTGGAGT 271

Qy          67 GGGTCTTTTTGTTG 80
                | ||||   ||||
Db         270 AGATCTTGGGGTTG 257


RESULT 7
AJ282728/c
LOCUS       AJ282728                 199 bp    mRNA    linear   EST 30-JUN-2000
DEFINITION  4A3A-ABB-D-05-R Anopheles gambiae immune competent 4A3A Anopheles
            gambiae cDNA clone 4A3A-ABB-D-05, mRNA sequence.
ACCESSION   AJ282728
VERSION     AJ282728.1  GI:6930607
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 199)
  AUTHORS   Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
            Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B.
            and Kafatos,F.C.
  TITLE     Anopheles gambiae pilot gene discovery project: identification of
            mosquito innate immunity genes from expressed sequence tags
            generated from immune-competent cell lines
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)
  MEDLINE   20300950
   PUBMED   10841561
COMMENT     Contact: Dimopoulos G
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES             Location/Qualifiers
     source          1..199
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="4A r/r"
                     /db_xref="taxon:7165"
                     /clone="4A3A-ABB-D-05"
                     /cell_line="immune competent 4A3A"
                     /lab_host="E. coli DH10B"
                     /clone_lib="Anopheles gambiae immune competent 4A3A"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from
                     forward priming site which reads from the 3' end of the
                     cDNA. The 4A3A is a directionally cloned and normalized
                     cDNA library that was constructed from the 4A3A cell line
                     oligo-T primed cDNA according to: Bonaldo, Lennon & Soares
                     (1996) : Normalization and Subtraction: Two approaches To
                     Facilitate Gene Discovery, Genome Research 6, 791-806."
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 1;  Length 199;
  Best Local Similarity   55.0%;  Pred. No. 28;
  Matches   60;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          12 TCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTC 71
               | | ||| |||     | ||| |||| |||  || |  ||    ||  |||  |  ||| 
Db         184 TGGGAGCAATTTAACGGCGCACTACTAACGGTAACGGACCAACACGTAATGTCCCGGGGT 125

Qy          72 TTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
                |||| || || | |  |     || ||  || ||  || | |  ||||
Db         124 CTTTTTTTTTGTTTCATTTGGCGGTAGTAAAATCGGCTTCTACTAGACT 76

RESULT 8
AV745983
LOCUS       AV745983                 677 bp    mRNA    linear   EST 18-OCT-2000
DEFINITION  AV745983 NPA Homo sapiens cDNA clone NPAAZG08 5', mRNA sequence.
ACCESSION   AV745983
VERSION     AV745983.1  GI:10865430
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Song,H., Peng,Y., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N.,
            Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S.,
            Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Ye,M.,
            Zhang,Q., Han,Z., Chen,Z., Hu,R. and Chen,J.
  TITLE     Homo sapiens NPA library cDNA clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Qinghua Zhang
            Shanghai Institute of Endocrinology, Rui-Jin Hospital
            197 Rui-Jin II Road, Shanghai 200025, P. R. China
            Tel: 86-21-64370045(ex.663332)
            Fax: 86-21-64743206
            Email: mbshi@ms.stn.sh.cn
            This clone is available at Shanghai Hematology Institute in
            Shanghai.
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi- Tech Park, Pudong.
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NPAAZG08"
                     /tissue_type="pituitary"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="NPA"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 1;  Length 677;
  Best Local Similarity   53.8%;  Pred. No. 34;
  Matches   63;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy           2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
               ||||  || || |         |  |   | |  | || ||||  || |    | |||| 
Db         436 GACCGCCTCCTGGATCTCCCCGCCGCAGCCTTCTCAGAAGACATCGAGCGGTCCTGAGAG 495

Qy          62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGA 118
                 || |||    |||||     ||||||| || ||  ||| |||| ||  ||   |||
Db         496 CCTCTGGGCACGTTTGTCTGTGTGCTGTAACCTGAAGTCAAACCTTAAGATAATGGA 552


RESULT 9
BM522137/c
LOCUS       BM522137                 494 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0674 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0807 5', mRNA sequence.
ACCESSION   BM522137
VERSION     BM522137.1  GI:18706191
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
```

```
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7506864|pir|T24294 hypothetical protein T01D3.3b -
            [Caenorhabditis elegans]
            Seq primer: T3 primer
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0807"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 494;
  Best Local Similarity   56.6%;  Pred. No. 44;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        135 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 76

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | |    |||  ||
Db         75 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 37


RESULT 10
BM522043/c
LOCUS       BM522043                 528 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0580 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0706 5', mRNA sequence.
ACCESSION   BM522043
VERSION     BM522043.1  GI:18705601
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7512096|pir|T17405 scavenger receptor cysteine-rich
            protein precursor - sea urchin [Strongylocentrotus purpuratus]
            Seq primer: T3 primer
            High quality sequence stop: 528.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0706"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 528;
  Best Local Similarity   56.6%;  Pred. No. 45;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        387 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 328

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | |    |||  ||
Db        327 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 289


RESULT 11
BM522233/c
LOCUS       BM522233                 539 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0771 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0908 5', mRNA sequence.
ACCESSION   BM522233
VERSION     BM522233.1  GI:18706995
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7512096|pir|T17405 scavenger receptor cysteine-rich
            protein precursor - sea urchin [Strongylocentrotus purpuratus]
            Seq primer: T3 primer
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0908"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
```

```
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 539;
  Best Local Similarity   56.6%;  Pred. No. 45;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        402 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 343

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        342 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 304


RESULT 12
BG817800/c
LOCUS       BG817800                 546 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  ESSU0222 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0293 5', mRNA sequence.
ACCESSION   BG817800
VERSION     BG817800.1  GI:14188780
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Seq primer: T3 primer
            High quality sequence stop: 546.
FEATURES             Location/Qualifiers
     source          1..546
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0293"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 546;
  Best Local Similarity   56.6%;  Pred. No. 45;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        407 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 348

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        347 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 309


RESULT 13
BM522287/c
LOCUS       BM522287                 572 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0825 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0963 5', mRNA sequence.
ACCESSION   BM522287
VERSION     BM522287.1  GI:18707474
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7512096|pir|T17405 scavenger receptor cysteine-rich
            protein precursor - sea urchin [Strongylocentrotus purpuratus]
            Seq primer: T3 primer
            High quality sequence stop: 572.
FEATURES             Location/Qualifiers
     source          1..572
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0963"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 572;
  Best Local Similarity   56.6%;  Pred. No. 45;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        425 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 366

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        365 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 327


RESULT 14
BG817916/c
LOCUS       BG817916                 577 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  ESSU0338 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0437 5', mRNA sequence.
ACCESSION   BG817916
VERSION     BG817916.1  GI:14188896
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
```

```
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to pir|T17405 scavenger recepto r cysteine-rich protein
            precursor - sea urchin (Strongylocen trotus purpuratus).
            Seq primer: T3 primer
            High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1..577
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0437"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 577;
  Best Local Similarity   56.6%;  Pred. No. 45;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        423 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 364

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        363 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 325


RESULT 15
BG817905/c
LOCUS       BG817905                 599 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  ESSU0327 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0423 5', mRNA sequence.
ACCESSION   BG817905
VERSION     BG817905.1  GI:14188885
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to pir|T17405 scavenger recepto r cysteine-rich protein
            precursor - sea urchin (Strongylocen trotus purpuratus).
            Seq primer: T3 primer
            High quality sequence stop: 599.
FEATURES             Location/Qualifiers
     source          1..599
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0423"
```

```
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 4;  Length 599;
  Best Local Similarity   56.6%;  Pred. No. 46;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        423 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 364

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        363 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 325


Search completed: August  6, 2005, 01:30:37
Job time : 1312.69 secs
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 21:21:24 ; Search time 1018.74 Seconds
                                           (without alignments)
                                           6135.749 Million cell updates/sec

Title:          US-10-782-899-1_COPY_472_600
Perfect score:  129
Sequence:       1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                    %
Result              Query
  No.     Score     Match Length DB  ID                     Description
 ----------------------------------------------------------------------------
     1      129     100.0    600  1  AB072446               AB072446 Bifidobac
     2      129     100.0    600  6  BD144660               BD144660 Drug for
 c   3      129     100.0  14216  1  AE014814               AE014814 Bifidobac
 c   4      129     100.0 156638  6  AX492788               AX492788 Sequence
 c   5      129     100.0 156646  6  AX553955               AX553955 Sequence
 c   6     32.6      25.3 295650  1  BX294151               BX294151 Pirellula
 c   7       32      24.8 194439  3  CNS07EGA               AL590443 chromosom
     8     31.8      24.7   3048  8  AK119844               AK119844 Oryza sat
 c   9     31.4      24.3  90417  2  AC138542               AC138542 Magnaport
 c  10     31.4      24.3  92770  2  AC138544               AC138544 Magnaport
    11     31.4      24.3 128428  2  AC144889               AC144889 Bos tauru
    12     31.4      24.3 207821  2  AC093397               AC093397 Bos tauru
    13       31      24.0   5832  1  AF105113               AF105113 Streptoco
    14       31      24.0  18754  1  AF094575               AF094575 Streptoco
 c  15     30.8      23.9 222330  2  AC096255               AC096255 Rattus no
 c  16     30.8      23.9 246369  2  AC125707               AC125707 Rattus no
    17     30.8      23.9 302050  1  AL935257               AL935257 Lactobaci
    18     30.6      23.7  26309  6  CQ363747               CQ363747 Sequence
    19     30.6      23.7 110000  1  AE017283_21            Continuation (22 o
    20     30.6      23.7 223086  2  AC130043               AC130043 Rattus no
 c  21       30      23.3    960  1  AY343890               AY343890 Streptomy
    22       30      23.3 236921  2  AC094151               AC094151 Rattus no
 c  23     29.8      23.1 100326 10   AL669928               AL669928 Mouse DNA
    24     29.8      23.1 151080  8  AP005726               AP005726 Oryza sat
 c  25     29.8      23.1 153065  2  AP005094               AP005094 Oryza sat
 c  26     29.6      22.9   2740  8  AK110305               AK110305 Oryza sat
 c  27     29.6      22.9  56276  7  AY129336               AY129336 Mycobacte
    28     29.6      22.9 110547  5  BX276099               BX276099 Zebrafish
 c  29     29.6      22.9 125020  9  AF429315               AF429315 Homo sapi
 c  30     29.6      22.9 164179  9  AC116901               AC116901 Homo sapi
 c  31     29.6      22.9 174118  9  AC024405               AC024405 Homo sapi
    32     29.6      22.9 185341  2  AC069403               AC069403 Homo sapi
    33     29.6      22.9 230664  2  AC097977               AC097977 Rattus no
    34     29.4      22.8    870  6  AR480420               AR480420 Sequence
    35     29.4      22.8   3520  6  CQ412044               CQ412044 Sequence
    36     29.4      22.8  16853  1  AF246897               AF246897 Streptoco
 c  37     29.4      22.8  37083  9  AC092021               AC092021 Homo sapi
    38     29.4      22.8 115469  9  AL356219               AL356219 Human DNA
    39     29.4      22.8 188183  9  AC105921               AC105921 Homo sapi
    40     29.4      22.8 188892 10   AL591946               AL591946 Mouse DNA
    41     29.4      22.8 203826 10   AC025669               AC025669 Mus muscu
    42     29.4      22.8 207136 10   AC122268               AC122268 Mus muscu
    43     29.4      22.8 301589  1  AE017312               AE017312 Desulfovi
 c  44     29.2      22.6    634  6  AX421454               AX421454 Sequence
    45     29.2      22.6   1265 10   BC026436               BC026436 Mus muscu

                                   ALIGNMENTS


RESULT 1
AB072446
LOCUS       AB072446                 600 bp    DNA     linear   BCT 15-JAN-2003
DEFINITION  Bifidobacterium longum biovar Longum hup gene for histone-like HU
            protein HB1, complete cds.
ACCESSION   AB072446
VERSION     AB072446.1  GI:19918953
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Takeuchi,A., Matsumura,H. and Kano,Y.
  TITLE     Cloning and expression in Escherichia coli of a gene, hup, encoding
            the histone-like protein HU of Bifidobacterium longum
  JOURNAL   Biosci. Biotechnol. Biochem. 66 (3), 598-603 (2002)
  MEDLINE   21999816
   PUBMED   12005055
REFERENCE   2  (bases 1 to 600)
  AUTHORS   Kano,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) Yasunobu Kano, Kyoto Pharmaceutical
            University, Institute of Molecular and Cellular Biology for
            Pharmaceutical Sciences; 1, Shichono-cho, Misasagi, Yamashina-ku,
            Kyoto 6078412, Japan (E-mail:ykano@mb.kyoto-phu.ac.jp,
            Tel:81755954719, Fax:81755954799)
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1679"
     RBS             180..185
     gene            193..474
                     /gene="hup"
     CDS             193..474
                     /gene="hup"
                     /codon_start=1
                     /transl_table=11
                     /product="histone-like HU protein HB1"
                     /protein_id="BAB88690.1"
                     /db_xref="GI:19918954"
```

```
                     /translation="MAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGL
                     KLTGLFSAERVKRPARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTE"
ORIGIN

  Query Match             100.0%;  Score 129;  DB 1;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 2
BD144660
LOCUS       BD144660                 600 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Drug for gene therapy using anaerobic bacteria.
ACCESSION   BD144660
VERSION     BD144660.1  GI:27850418
KEYWORDS    JP 2002097144-A/1.
SOURCE      Bifidobacterium longum
  ORGANISM  Bifidobacterium longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Fujimori,M., Taniguchi,S., Amano,J., Yazawa,K., Kano,Y.,
            Nakamura,T. and Sasaki,T.
  TITLE     Drug for gene therapy using anaerobic bacteria
  JOURNAL   Patent: JP 2002097144-A 1 02-APR-2002;
            JUN AMANO,MINORU FUJIMORI
COMMENT     OS   Bifidobacterium longum
            PN   JP 2002097144-A/1
            PD   02-APR-2002
            PF   21-SEP-2001 JP 2001290187
            PI   MINORU FUJIMORI,SHUNICHIRO TANIGUCHI,JUN
            AMANO,KAZUTORA YAZAWA,
            PI   YASUMASA KANO,TOSHIYUKI NAKAMURA,TAKAYUKI SASAKI PC
            A61K35/74,A61K31/396,A61K31/7068,A61K31/711,A61K38/00,A61K38/ PC
            43,A61K48/00,
            PC   A61P35/00,C12N1/21,C12N15/09//(C12N1/21,C12R1:01),C12N15/00,
            PC   A61K37/02,
            PC   A61K37/48
            CC   Drug for gene therapy using anaerobic bacteria FH   Key
                 Location/Qualifiers
            FT   CDS                 (193)..(471).
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:216816"
ORIGIN

  Query Match             100.0%;  Score 129;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
```

```
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 3
AE014814/c
LOCUS       AE014814                14216 bp    DNA     linear   BCT 31-OCT-2002
DEFINITION  Bifidobacterium longum NCC2705 section 201 of 202 of the complete
            genome.
ACCESSION   AE014814 AE014295
VERSION     AE014814.1  GI:23327097
KEYWORDS    .
SOURCE      Bifidobacterium longum NCC2705
  ORGANISM  Bifidobacterium longum NCC2705
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1  (bases 1 to 14216)
  AUTHORS   Schell,M.A., Karmirantzou,M., Snel,B., Vilanova,D., Berger,B.,
            Pessi,G., Zwahlen,M.-C., Desiere,F., Bork,P., Delley,M.,
            Pridmore,D. and Arigoni,F.
  TITLE     The genome sequence of Bifidobacterium longum reflects its
            adaptation to the human gastrointestinal tract
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14422-14427 (2002)
   PUBMED   12381787
REFERENCE   2  (bases 1 to 14216)
  AUTHORS   Schell,M.A., Karmirantzou,M., Snel,B., Vilanova,D., Berger,B.,
            Pessi,G., Zwahlen,M.-C., Desiere,F., Bork,P., Delley,M.,
            Pridmore,D. and Arigoni,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) Bioscience, Nestle Research Center, P.O.
            Box 44, Lausanne 26 1000, Switzerland
FEATURES             Location/Qualifiers
     source          1..14216
                     /organism="Bifidobacterium longum NCC2705"
                     /mol_type="genomic DNA"
                     /strain="NCC2705"
                     /db_xref="taxon:206672"
     gene            178..1638
                     /gene="BL1797"
     CDS             178..1638
                     /gene="BL1797"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical proteasome-associated protein"
                     /protein_id="AAN25580.1"
                     /db_xref="GI:23327098"
                     /translation="MPQLRDSGTRSLHATEPVPSAEMDGFCRIFGVETEYGVAVTGAE
                     RPVDAGQVAMTMFQPIVSRSRSTNTYLANGSRLYLDVGSHPEYATAEARDPREALAQD
                     LAGEHVMRNLALKAQRKLRESYGAHATIHVFKNNVDSAGHAFGCHENYLVRRFVPLET
                     IEHQLLPFLITRQLYTGAGRMTPDGFQITQRADFLDEAVSSATTRSRPMVNTRDEPHA
                     DPDSFRRLHVIIGDSNRSQWSTWMKLAVTHLVLCAIEDAFRHGVPSGFEQYAFADPAA
                     ANRTVSRFLDNPRAELTLESGESVSALGLQRRYYAAVKAFIETHGDALAGSLPATTID
                     TIMGEWSRVLDALERGAYDALADRVDWAAKKCLFDALKRRRPDVTFAQMEQLELDYHD
                     IANGRLYGSLVARNQMRELLTGDNVEYAVHNPPTDTRAALRGRFVDAALNVGAQFSAD
                     WTHLTLTAPERREAILLDPFEAEPTLGFEQLMEALN"
     gene            complement(1754..2035)
                     /gene="hup"
                     /note="synonym: BL1798"
     CDS             complement(1754..2035)
                     /gene="hup"
                     /note="HS1; Hup; COG family: bacterial nucleoid
                     DNA-binding protein; PFAM_ID: Bac_DNA_binding"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-binding protein Hu"
                     /protein_id="AAN25581.1"
                     /db_xref="GI:23327099"
                     /translation="MAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGL
                     KLTGLFSAERVKRAARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTE"
     gene            complement(2181..4742)
                     /gene="BL1799"
     CDS             complement(2181..4742)
                     /gene="BL1799"
```

```
                     /note="COG family: predicted integral membrane protein"
                     /codon_start=1
                     /transl_table=11
                     /product="narrowly conserved hypothetical membrane
                     protein"
                     /protein_id="AAN25582.1"
                     /db_xref="GI:23327100"
                     /translation="MTTEQDQSTTEDRPDKATSSTTDRPIASGDGRSASFAVSGPDSP
                     ATDDSNDSGPYIDDIAPRRTRDFGDLTRAGLSLLMAAVVMVFAVYLGGMTRGVESDAH
                     TAAQVINWLADFPSTVLTQLATIVIVIIVLAQLLLAREWLQAAVSALAMFAGYGMVWV
                     VSTAISSLNDFTLPMALVSAATSYGSGLLPDIYAGMASFLTAAGPRRTRSTVKWSWNI
                     LYAIAAVMVVLSWHSVTGMLVSMAAGRTVGMLIRFVVGTQNKGVWGQDLVAALKGIGL
                     DTASLIRHQEPGVGHGSLAVTLDDDLAEGSRIYDLDTTDNRRFIVSVIDAQTHTVGYL
                     KQLWDWVRFTSVSIRRDRSVRDAVQHHFAMLLGLHNIKLPAPAPYGIADTDESAILVL
                     DAHTIELPANLNTLTKADAVAYMRYLSVASRRGYTHRRITPDTLARLEDGTAVIAGWL
                     NGDNASGPANTALDKVQLLVLFAALIGVEPAVEAAREAWGDTTLTALAPFIQKVAVPS
                     PTRALESWDKQLLKGLRSRITALADEETVESAEPVTLARFSWRSMITMLLVIVAVVVV
                     FTQLKPEEIITALTNANPLMAVVTLAFGVCGWIGSSISLGSLMARHKRDNMGVFMSQV
                     AGGFATVSMPAGVGPSFVNLQFLRKSGYRNTQATAIMSAALVVYYAVYFSMLVIIGLF
                     TGRNMFSGAIPTNTLVIVLGVVVVVLSIAMMIPPLRHWVTRRLMPLAKTYINQLLDVL
                     SQPRQLTVSCLGALFQNATTGLAFWAALQAFGHSSNPIETTFVFLLAYALGSAVPTPG
                     GLGGVEAALTFAFVAVGVPQGVALSATLLHRVVFYWLRIPLGAAAMKWLDKHNLV"
     gene            complement(4874..6319)
                     /gene="purB"
                     /note="synonym: BL1800"
     CDS             complement(4874..6319)
                     /gene="purB"
                     /note="adenylosuccinase; PurB; COG family:
                     adenylosuccinate lyase; PFAM_ID: lyase_1"
                     /codon_start=1
                     /transl_table=11
                     /product="adenylosuccinate lyase"
                     /protein_id="AAN25583.1"
                     /db_xref="GI:23327101"
                     /translation="MEVMKLTDISPAIALTPLDGRYHKATAPLVEYMSEPALNRERMR
                     VEVEWMILLANGFEGNGNQPIVDGVKPFTAEEQAFLRAIPEDFGAEGIKQHAAHEAIT
                     HHDVKAVEYYIDDQIDKAPAELTNINDELKTLVHFACTSEDINNLSIARCVKNAMENV
                     WTPEFKRIIDHLAGKAEEFKDMPLLSLTHGQPATPTTLGKELAVYVYRLNRQLKHIEN
                     QEYLGKINGATGTFGAHLAACPDVDWIAVSREFVTNRMGLTWNPLTTQIESHDWQAEL
                     YSTVSHANRIMHNLCVDVWMYISRGVFAQVPVKGATGSSTMPHKVNPIRFENAEANFE
                     ISCSLLDTLSATLVESRWQRDLTDSTTQRNIGSALGYSLLALTNLMGGLESIHPNTHV
                     IERELDENWEVLGEPIQTAMRACELKGLPGMDKPYEKVKGLMRGRTINKEQVEQFIDQ
                     QSFDPETAARLKALTPATYTGVASQLVAFGR"
     gene            6443..7084
                     /gene="BL1801"
     CDS             6443..7084
                     /gene="BL1801"
                     /note="COG family: permeases of the major facilitator
                     superfamily"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical membrane protein with unknown
                     function"
                     /protein_id="AAN25584.1"
                     /db_xref="GI:23327102"
                     /translation="MPATTIHTPPKRPIAVSSLKWQPLDIAVGATLGAACGVIFWGFN
                     FAYMPISSVFGAILPGFASIVHAVWYFSGTLAVLILRKPGAALYVNLIGTLVETVIGS
                     QFDLGFVIISALLQGVFAEIPFALTRYRMFNLPISIASGALVALEYGTYLMLFRYVGV
                     SFVSPRGIVHMISEVTGGVLISGVMAWYLMLAIAKTGVLDRLASGRALIERRA"
     gene            complement(7220..7585)
                     /gene="BL1802"
     CDS             complement(7220..7585)
                     /gene="BL1802"
                     /note="COG family: DNA-directed RNA polymerase specialized
                     sigma subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25585.1"
                     /db_xref="GI:23327103"
                     /translation="MKTYQVIVRPGEKYWILEIPGIGFTQARTTAEIESMARDLITVM
                     TQDADFALTVETKLPQSVQEHLDEARRLRKEEAECRSNAARETRAAAQELHGMGLALQ
                     QIGDILGVSRQRAHQLVNA"
```

```
     gene            7610..7855
                     /gene="BL1803"
     CDS             7610..7855
                     /gene="BL1803"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25586.1"
                     /db_xref="GI:23327104"
                     /translation="MASASLQIMSARLLSISLCRGMGNHKSVELEYLMTPSLPDTFKR
                     DILAFCRLLHQFDEPSPTRHYIVNSFLDNSVKCGIYF"
     gene            complement(7852..9468)
                     /gene="BL1804"
     CDS             complement(7852..9468)
                     /gene="BL1804"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25587.1"
                     /db_xref="GI:23327105"
                     /translation="MGVVMSEHDYWQVESSVYGGVGYAPATLEEYAAIAKALDDEAAG
                     FAAIATAWESAALQLQSHRHSAPMCVTLQSGDPSAVVPGHVTAPYAALGNRCYDHATA
                     CQRLSDDLRGAADLLIRAHSLYSQAEMTARRMFTELLQAGTQAKPGYAAAGVAAVAAG
                     GFLAGWTIDGKPNPAWMSTFTYPFQEGVLSGAGGIIGGVPIGKSIAHTDEVNKAAGKI
                     ANFSGPAKDVVQGNHLDVREVQANADVVRASGSVAESMENLRRLAEERLGKIELNSGL
                     EYGTIAIQRYERSDGTNSWLVTIPGTDGQPDSPFGWAQNVELMSADQERRRKADSARM
                     VAEAMRQAGIGKDEPVALIGHSQGGIVAATLASDWAEEYTIEHVVTAGSPVANHPIPQ
                     RTWVTSVEIDDELVAALDGAANPVTDNWLTVQGHVSPAPAATPSTVHSDGSCTPGATP
                     ITGLTPYDAAPVAGSTNGRELSHWIKYHQAAYQNATDLGSPAVQRHEAHFQEVINGEL
                     KETRYYQGRMTQSATIAPSERTTEFSTFGG"
     gene            complement(9453..9800)
                     /gene="BL1805"
     CDS             complement(9453..9800)
                     /gene="BL1805"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25588.1"
                     /db_xref="GI:23327106"
                     /translation="MSLVLLAGYMNTYTSNPAPIDPTGGFAGNTVTPFDYGTCPCHAA
                     VQRSRDLVSVQRAATLAQSDQTAVAVNRSSNIVWTGKAADLFRSRLSRTAQYSASLAQ
                     DMETTHRLAWGSS"
     gene            complement(9933..10583)
                     /gene="BL1806"
     CDS             complement(9933..10583)
                     /gene="BL1806"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25589.1"
                     /db_xref="GI:23327107"
                     /translation="MSALGTYRPGLWRRIAVAGCRRGGSKPVRFDYVQSGHRIIERQY
                     QSGSGRVHGYHHAQGATAADRRAIRRGRANAHASVAASQGRHRHQCFDFFRADKDHPE
                     TGRGAEQRFRFRASPVGTAIREFFIQCQKGRRVDRRTEEAGRGADEGQSAIHRYPVEH
                     HPIRAESDAEQGYRGYQAVVSSRSQLFFFLTPSRVMALSRLFPGFGCRLWHLSNFR"
     gene            complement(10642..11673)
                     /gene="BL1807"
     CDS             complement(10642..11673)
                     /gene="BL1807"
                     /note="COG family: uncharacterized Bcr"

  Query Match             100.0%;  Score 129;  DB 1;  Length 14216;
  Best Local Similarity   100.0%;  Pred. No. 2e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 1697

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1696 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 1637
```

```
Qy        121 AGTTCAGCG 129
              |||||||||
Db       1636 AGTTCAGCG 1628


RESULT 4
AX492788/c
LOCUS       AX492788              156638 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 1106 from Patent EP1227152.
ACCESSION   AX492788
VERSION     AX492788.1  GI:23338471
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   None.
  TITLE     Bacterial strain and genome of bifidobacterium
  JOURNAL   Patent: EP 1227152-A 1106 31-JUL-2002;
            Societe des Produits Nestle S.A. (CH)
FEATURES             Location/Qualifiers
     source          1. .156638
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="1106 seq in place of 1099 because seq 1 (2.256.638
                     bases splitted-in 7 more sequences.~seq 0001: from
                     0.000.001 to 0.349.980 length: 349980~seq 1100: from
                     0.300.001 to 0.649.980 length: 349980~seq 1101: from
                     0.600.001 to 0.949.980 length: 349980~seq 1102: from
                     0.900.001 to 1.249.980 length: 349980~seq 1103: from
                     1.200.001 to 1.549.980 length: 349980~seq 1104: from
                     1.500.001 to 1.849.980 length: 349980~seq 1105: from
                     1.800.001 to 2.149.980 length: 349980~seq 1106: from
                     2.100.001 to 2.256.638 length: 156638"
ORIGIN

  Query Match            100.0%;  Score 129;  DB 6;  Length 156638;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141691 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 141632

Qy         61 TGGTCGGGGTCTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141631 TGGTCGGGGTCTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 141572

Qy        121 AGTTCAGCG 129
              |||||||||
Db     141571 AGTTCAGCG 141563


RESULT 5
AX553955/c
LOCUS       AX553955              156646 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 9 from Patent WO02074798.
ACCESSION   AX553955
VERSION     AX553955.1  GI:25897908
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Arigoni,F., Delley,M., Mollet,B., Pridmore,R.D., Schell,M.A.,
            Pohl,T.G. and Zwahlen,M.C.
  TITLE     The genome of a bifidobacterium
  JOURNAL   Patent: WO 02074798-A 9 26-SEP-2002;
            SOCIETE DES PRODUITS NESTLE S.A. (CH)
FEATURES             Location/Qualifiers
     source          1. .156646
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="seq 1: original length 2256646 splitted into~seq 1:
                     from 0.000.001 to 0.349.980~seq 3: from 0.300.001 to
                     0.649.980~seq 4: from 0.600.001 to 0.949.980~seq 5: from
                     0.900.001 to 1.249.980~seq 6: from 1.200.001 to
                     1.549.980~seq 7: from 1.500.001 to 1.849.980~seq 8: from
                     1.800.001 to 2.149.980~seq 9: from 2.100.001 to 2.256.646"
ORIGIN

  Query Match            100.0%;  Score 129;  DB 6;  Length 156646;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141699 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 141640

Qy         61 TGGTCGGGGTCTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141639 TGGTCGGGGTCTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 141580

Qy        121 AGTTCAGCG 129
              |||||||||
Db     141579 AGTTCAGCG 141571


RESULT 6
BX294151/c
LOCUS       BX294151              295650 bp    DNA     linear   BCT 11-JUL-2003
DEFINITION  Pirellula sp. strain 1 complete genome; segment 19/24.
ACCESSION   BX294151 BX119912
VERSION     BX294151.1  GI:32446812
KEYWORDS    complete genome.
SOURCE      Rhodopirellula baltica SH 1
  ORGANISM  Rhodopirellula baltica SH 1
            Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
            Planctomycetaceae; Rhodopirellula.
REFERENCE   1  (bases 1 to 295650)
  AUTHORS   Gloeckner,F.O., Kube,M., Bauer,M., Teeling,H., Lombardot,T.,
            Ludwig,W., Gade,D., Beck,A., Borzym,K., Heitmann,K., Rabus,R.,
            Schlesner,H., Amann,R. and Reinhardt,R.
  TITLE     Complete genome sequence of the marine planctomycete Pirellula sp.
            strain 1
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)
  MEDLINE   22735913
   PUBMED   12835416
REFERENCE   2  (bases 1 to 295650)
  AUTHORS   Kube,M., Borzym,K., Heitmann,K., Klages,S., Marquardt,I.,
            Lehrack,S., Beck,A., Pawlik,R., Reinhardt,R., Gloeckner,F.O.,
            Bauer,M., Teeling,H., Lombardot,T., Ludwig,W., Gade,D., Rabus,R.,
            Schlesner,H. and Amann,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2003) Max Planck Institute for Molecular
            Genetics, proScience Ihnestrasse 73, D-14195 Berlin, Germany Max
            Planck Institute for Marine Microbiology Celsiusstrasse 1, D-28359
            Bremen, Germany
COMMENT     This project was carried out by
            *Max Planck Institute for Molecular Genetics, Berlin, Germany; *Max
            Planck Institute for Marine Microbiology, Bremen, Germany; in the
            fremework of the REGX-project, http://www.regx.de --------------
            Genome Center
                 Center: Max Planck Institute for Molecular Genetics
                 Center code: MPIMG
            -------------- Summary Statistics
                 Sequencing vector: pUC19; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 7142841 bases at least Q40
                 Consensus quality: 7145138 bases at least Q30
                 Consensus quality: 7145484 bases at least Q20
```

```
            Quality coverage: 8.03
         --------------
         This sequence was finished as follows unless otherwise noted: all
         regions were double stranded, sequenced with an alternate
         chemistry, or coveredby high quality data (i.e., phred quality >=
         30); an attempt was made to resolve all sequencing problems, such
         as compressions and repeats; allregions were covered by at least
         one plasmid Sequence; assembly was additionally confirmed by long
         range pcr and cosmid end sequences.
         --------------
         See http://www.micro-genomes.mpg.de/pirellula/ for more information
         including minimal tiling path from a set of  220 cosmids out of
         908. See the misc_feature tag below for the boundaries of the MTP
         cosmids. -------------- Annotation
            Center: Max Planck Institute for Marine Microbiology
                    Celsiusstrasse 1, D-28359 Bremen, Germany.
            Center Code: MPIMM
            Email: fog@mpi-bremen.de
            Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 580
         --------------
         Three different programs (Glimmer, Critica and Orpheus) were used
         for ORF-prediction. A nonredundant list of ORFs was generated by
         suitable parsing of the results.
         Automated annotation was done with the software package Pedant Pro
         (http://www.biomax.de). All ORF predictions and annotations were
         manually corrected by considering all results of the different
         tools applied. See http://www.regx.de for more information and
         access to supplementary information.
         --------------.
FEATURES             Location/Qualifiers
     source          1. .295650
                     /organism="Rhodopirellula baltica SH 1"
                     /mol_type="genomic DNA"
                     /strain="1"
                     /db_xref="taxon:243090"
     gene            complement(414. .1508)
                     /gene="pknH"
                     /locus_tag="RB10133"
     CDS             complement(414. .1508)
                     /gene="pknH"
                     /locus_tag="RB10133"
                     /EC_number="2.7.1.-"
                     /function="unspecified signal transduction; unspecified
                     kinase or ATP dependent regulatory protein; cellular
                     communication/signal transduction"
                     /note="best DB hits: BLAST: swissprot:Q11053; PKNH_MYCTU
                     PROBABLE SERINETHREONINE-PROTEIN; E=3e-35 embl:CAB94054.1;
                     (AL358672) putative serinethreonine-protein; E=1e-34
                     embl:CAA10713.1; (AJ132604) hypothetical protein
                     [Lactococcus; E=2e-33 COG: Rv1266c; COG0515
                     Serine/threonine protein kinases; E=3e-36 PFAM: PF00069;
                     Protein kinase domain; E=7.6e-54"
                     /codon_start=1
                     /transl_table=11
                     /product="probable serine/threonine-protein kinase pknH"
                     /protein_id="CAD78719.1"
                     /db_xref="GI:32446813"
                     /translation="MMPHLRLTDFELGAVLGVGTVGTVYDGKILDDVEVHPAAEAIRG
                     QDLAVKKLHPAVSQDDLIQARFRREMVILERLQHPNIIGYFGGGSEDGQLFYVMERVD
                     GGTIKDLLETNGALAWPVVVDVARQVCSALQCAHNHGVIHRDLKPGNLFLTRDAHVKL
                     GDFGIARDQHSSDLTSQGLTVGTHAYMAPEQITGDEAISGKADLYALGCVLFEMLANR
                     KVFAGENFAQLFEQHLRTKAPTIASIVPDVPPELSQVIADCLEKSPDDRPFNARSVQG
                     VMIEIGEKYNLSAPPDSPVGPNHSETGNAPKDVSADSVTEKGRRLLEEQIHYRLGGTS
                     REPVSAGKVSLIVVAIIVLIVLAVSLSGGS"
     gene            complement(1529. .1843)
                     /locus_tag="RB10134"
     CDS             complement(1529. .1843)
                     /locus_tag="RB10134"
                     /function="ribosomal proteins"
                     /note="best DB hits: BLAST: gb:AAK05177.1; AE006340_10
                     (AE006340) 50S ribosomal protein L21; E=4e-10 pir:F82920;
                     ribosomal protein L21 UU212 [imported] - Ureaplasma;
                     E=3e-11 gb:AAC35614.1; (AF041468) ribosomal protein L21
                     [Guillardia; E=3e-10 COG: Rv2442c; COG0261 Ribosomal
                     protein L21; E=3e-07 PFAM: PF00829; Ribosomal prokaryotic
                     L21 prot; E=2.2e-36"
                     /codon_start=1
                     /transl_table=11
                     /product="50S ribosomal protein L21"
                     /protein_id="CAD78720.1"
                     /db_xref="GI:32446814"
                     /translation="MYAIFVDGGRQYRVEPGMELDVDYRDIAAGENLKFETVLAVGAE
                     DGLKLGAPTLDGVSVSASVLGMSQDKKIYIQKFRRRKHSKKRTGHRQKNTRIRIEEIA
                     GV"
     gene            complement(1767. .2018)
                     /locus_tag="RB10135"
     CDS             complement(1767. .2018)
                     /locus_tag="RB10135"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD78721.1"
                     /db_xref="GI:32446815"
                     /translation="MLDRPIFGSTDLFTSDDLLDSLALKNQPWSKRRSVVILRDSKIR
                     PSGQFPLHPEPFSRHVCHLCRRWTPIPSRARNGTRRRLP"
     gene            2019. .2177
                     /locus_tag="RB10136"
     CDS             2019. .2177
                     /locus_tag="RB10136"
                    \/codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD78722.1"
                     /db_xref="GI:32446816"
                     /translation="MLVLRNRSELSVGYSGAARAARCVTGGLAPNGTYFTPPQGGCEF
                     LRASPVTP"
     gene            complement(2151. .2372)
                     /locus_tag="RB10137"
     CDS             complement(2151. .2372)
                     /locus_tag="RB10137"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein-putative conserved
                     hypothetical protein"
                     /protein_id="CAD78723.1"
                     /db_xref="GI:32446817"
                     /translation="MSRSRFSGVLYDAPLPKRSLGGGRATPFRRTRGRGPSMENGADC
                     PPPEISLNARFPTPPAARAGFSRRYWRCT"
     gene            complement(2478. .3281)
                     /locus_tag="RB10138"
     CDS             complement(2478. .3281)
                     /locus_tag="RB10138"
                     /function="inner membrane organization"
                     /note="best DB hits: BLAST: pir:T34929; hypothetical
                     protein SC3F9.07 SC3F9.07 - Streptomyces; E=0.018"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein-putative membrane
                     lipoprotein"
                     /protein_id="CAD78724.1"
                     /db_xref="GI:32446818"
                     /translation="MRLVTDFFNAFPFLQRNSTSSALMSLVFAGCCLLNVGFESIAEA
                     DEPAEQTSNVSVKIFDTNAPGWRQLDGDDFVRVNGDDQTLTWEGTEALGSGQPIGVTR
                     TNFEVKNFELVIQWKHLKPAGNSGVFAWVPMSALKDLPPNRLPNTGIEVQMLDLDYGR
                     MYTEKTGKPPTWFTSHGDIFAVGKSSMQPFPPLSPDGHRSFPSAETTNPHGEWNQYYV
                     RGINREIRLWVNGVEVSGGRSCSPDEGFLCLESEGSPIRFREIWLRELP"
     gene            complement(3382. .3621)
                     /locus_tag="RB10139"
     CDS             complement(3382. .3621)
                     /locus_tag="RB10139"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD78725.1"
                     /db_xref="GI:32446819"
                     /translation="MRHKHDRMKIRGISKNPIDLPSRYRGRVLKPKKRRRSFRSFLRE
```

```
                 SRTIILSYHGSTRIHLAQPAFGRCHNPQSSTHQKS"
     gene        3707. .5527
                 /locus_tag="RB10140"
     CDS         3707. .5527
                 /locus_tag="RB10140"
                 /function="anion transporters (Cl(-), SO4(2-), PO4(3-),
                 etc.)"
                 /note="best DB hits: BLAST: pir:B82127; sulfate permease
                 family protein VC2031 [imported] -; E=1e-89 pir:G83325;
                 probable sulfate transporter PA2563 [imported] -; E=5e-53
                 ddbj:BAB04133.1; (AP001508) sulfate permease [Bacillus;
                 E=1e-51 COG: VC2031; COG0659 Sulfate permease and related

  Query Match             25.3%;  Score 32.6;  DB 1;  Length 295650;
  Best Local Similarity   53.5%;  Pred. No. 18;
  Matches   68;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              |||| |    |||| | ||| || ||||| |      |||||| ||||||     | |||
Db     148951 TGACGTGAGCCTCGCATCGAATAGTTCGAACGACGACGACGACGAAGACCAACGTCAAGA 148892

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              || |   | |  ||| |  | | | | |   |     |    | | |  | | ||| |||
Db     148891 TGTTGTCGTTGATTTGGCGGAGTTTCGGATGCAGTCCGAAAGAGCATTCTCTACCGAACT 148832

Qy        121 AGTTCAG 127
                ||| |
Db     148831 GTTTCGG 148825


RESULT 7
CNS07EGA/c
LOCUS       CNS07EGA              194439 bp    DNA     linear   INV 30-JUL-2004
DEFINITION  chromosome III of strain GB-M1 of Encephalitozoon cuniculi
            (Microspora).
ACCESSION   AL590443
VERSION     AL590443.1  GI:13560062
KEYWORDS    AIDS opportunistic pathogen; Genome sequence; intracellular
            parasite.
SOURCE      Encephalitozoon cuniculi GB-M1
  ORGANISM  Encephalitozoon cuniculi GB-M1
            Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
REFERENCE   1  (bases 1 to 194439)
  AUTHORS   Katinka,M.D., Duprat,S., Cornillot,E., Metenier,G., Thomarat,F.,
            Prensier,G., Barbe,V., Peyretaillade,E., Brottier,P., Wincker,P.,
            Delbac,F., El Alaoui,H., Peyret,P., Saurin,W., Gouy,M.,
            Weissenbach,J. and Vivares,C.P.
  TITLE     Genome sequence and gene compaction of the eukaryote parasite
            Encephalitozoon cuniculi
  JOURNAL   Nature 414 (6862), 450-453 (2001)
   MEDLINE  21576510
   PUBMED   11719806
REFERENCE   2  (bases 1 to 194439)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     Corresponding author : Equipe de Parasitologie Moleculaire et
            Cellulaire, UMR CNRS 6023, Universite Blaise Pascal, 63177 Aubiere
            cedex, France E-mail : Christian.VIVARES@lbp.univ-bpclermont.fr
            Tel : (33) (0)4.73.40.74.57
            Fax : (33) (0)4.73.40.74.55
            Web : http://www.protistes.univ-bpclermont.fr/fr/parasito.htm The
            submitted sequence encompasses the coding core of the chromosome.
            The telomeric and subtelomeric sequences are not included in this
            submission.
FEATURES             Location/Qualifiers
     source          1. .194439
                     /organism="Encephalitozoon cuniculi GB-M1"
                     /mol_type="genomic DNA"
                     /strain="GB-M1"
                     /db_xref="taxon:284813"
                     /chromosome="III"
     gene            550. .1380
                     /gene="ECU03_0010"
     CDS             550. .1380
                     /gene="ECU03_0010"
                     /function="unknown"
                     /note="03_0010, unknown, gene found by Glimmer [Bad Olap 2
                     167 0]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26148.1"
                     /db_xref="GI:19168629"
                     /translation="MAAPTQLPETAKPKHSNQSEAGPAPLASPTAPMPRPASHLAPMP
                     SDHPDFRSKSARLRCQPPRTNNCGTFKQPPSVAATSRPKPGNPFLQPPTKGTPPPKKK
                     KKNHTEGCHTHEANPEPNTKHTETEPPIISHCPPPHPGPTATPNLLPCPNPTSSFCQN
                     TRDSPSLPPQRPNMVHIPETPLQRRHSPSQIPIRIPPSPYNLSTAKTYYYLQPPYLPI
                     LSHEIHFIVPLLCPYEHHTHNRAARREPATLENNPRHRSIRLNYIPSHHVPHLRRRQF
                     "
     gene            1211. .2020
                     /gene="ECU03_0020"
     CDS             1211. .2020
                     /gene="ECU03_0020"
                     /function="unknown"
                     /note="03_0020, unknown, gene found by Glimmer [Shorter 1
                     167 0]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26149.1"
                     /db_xref="GI:19168630"
                     /translation="MKSISSYLSFAHMNITHTTEQHAENQPHWKTILDIAPFVSITFP
                     AIMCLIFDEDSFEESPFLRFITLLLPFSYSAVQYALLYTNWKSHNKPEPILHTTLYYT
                     LSLLLLAFTIISILSIIPFSLNEWDHAASFFYPIVLPSFTVPPAYLLSSSYFLVPRQI
                     RLTDTVISILISVCSIVNVLLVFKEFNYYPYSAIISSISVLLQLLSEKHCLFKQSPPS
                     TASSRAAVLILTLILAVLVYTFLGYGAIYILDDHFHLLGKMKSILPSEPHQ"
     gene            complement(3222. .3578)
                     /gene="ECU03_0030"
     CDS             complement(3222. .3578)
                     /gene="ECU03_0030"
                     /function="unknown"
                     /note="03_0030, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26150.1"
                     /db_xref="GI:19168631"
                     /translation="MCFPVGGLLPYLALRSVAYICVFGDSSRAPDLIREAFESPHFKR
                     FDGAGTYKEAKGRCGMRFADVVNGAFGQISDMADKVGKGEPEVWCIWKKRGEVEMLLK
                     VKEYRKGYGSGKRRRR"
     gene            4169. .4975
                     /gene="ECU03_0040"
     CDS             4169. .4975
                     /gene="ECU03_0040"
                     /function="unknown"
                     /note="03_0040, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26151.1"
                     /db_xref="GI:19168632"
                     /translation="MNTTHVPEPHRTEQHTENLQHWRKILGIAPIVSIAFPAIMYFIF
                     TKDSFEDSLFLRFITILLSFSYSAVQYAVLLHTNWKSHNKPEGILHTTLYYTLNLLLL
                     AFSIISILSITTLPINKWKNDGGPILFSIFLPPLFMSPAYLLSTSCRLVPGQIGFTDT
                     GINVLIDILTLLCSVGSLLLILEESEYCYCFAIISSILILIRLLGEKLSPEKQSPPPT
                     APWRIAVFVLILIFAALIYAFMMWVSIDILSDHFDLLTKARSTSVSKPGQ"
     gene            5400. .6665
                     /gene="ECU03_0050"
     CDS             5400. .6665
                     /gene="ECU03_0050"
                     /function="unknown"
                     /note="03_0050, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26152.1"
                     /db_xref="GI:19168633"
                     /translation="MEQISRIKTTLDEVDLDYDERAFRGLSVEKLWVRAFWLISDIVK
```

```
                     CEKMHPAIACYRNLLDSILEDLEEKGHPEGITAVNAAKDALRALEDVVNRKCQILSLS
                     KKAHGTEFPYYFGAIDGYLKDYPGLLKAFKDALRKVLPHVRGDLKVKVNEMKLEAAMS
                     DFLKVLKKQLDGEKVQERPSTLDIYKYLMNEVNEVEVPLKSVIRDLLGIEGNYNFEAN
                     WSFYSYPRDINEAALYIFQKYFQHDIDEASLSRTIKNELECLEKVISKVPKSWRARLQ
                     RILRDVRTAFDSLSQKLKALRTSCTRVIDDVESNGKSFSDAFSELNGMLREAEKGLSD
                     LKERISDLERVLPSSPPMKLCELLCFMFIAILAIHSFLHLAEDKNSRVKTRWNTALYS
                     LAVAGTIMHQMWVLWGGGGTGMRMGQVMC"
     gene            6782. .7153
                     /gene="ECU03_0060"
     CDS             6782. .7153
                     /gene="ECU03_0060"
                     /function="unknown"
                     /note="03_0060, unknown, gene found by Glimmer [Delay by
                     228 5 109 38]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26153.1"
                     /db_xref="GI:19168634"
                     /translation="MCGQRLIQCGMSWKQMCVFCAGHVLLAVACFAEPSKHPLLDTYG
                     RCGMVCIGLTVLIALLLLVSYMDEADGEKASASGSGAVVMNCLFVMSVAVTTFASWVS
                     SRDYFMTCRRGTDDMTGTAAS"
     gene            complement(7642. .7953)
                     /gene="ECU03_0070"
     CDS             complement(7642. .7953)
                     /gene="ECU03_0070"
                     /function="unknown"
                     /note="03_0070, unknown, gene found by Glimmer [Shorter 8
                     242 99]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26154.1"
                     /db_xref="GI:19168635"
                     /translation="MIERERETVLYFQLWMANAMTERMEMTANVMKRALEKYRSILCM
                     APSGLVLSFQLARNRRTIGNAIDMNTARSLGIVPPPIIELVLGVLLCCVCGVHAGKGE
                     V"
     gene            7712. .8389
                     /gene="ECU03_0080"
     CDS             7712. .8389
                     /gene="ECU03_0080"
                     /function="unknown"
                     /note="03_0080, unknown, gene found by Glimmer [Bad Olap 7
                     242 0]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26155.1"
                     /db_xref="GI:19168636"
                     /translation="MGGGTIPRDLAVFMSIAFPMVLLFRASWKESTKPEGAIHKILRY
                     FSSALFITFAVISILSVIAFAIHNWKYNTVSLSLSIIFPSLLFPPTYLLSTSCCLVPG
                     QIGFTDTGINILIDILILVHPMANLVSVRRRPYRYPHSAIISSIFLIIIFTPDNLNNG
                     ATAFCSQLSPPYLDLQPIFSVLSASRQHHIPAFLLALACPIMGLIPMIKDSSTIPISR
                     LPLLPSF"
     gene            complement(8576. .10444)
                     /gene="ECU03_0090"
     CDS             complement(8576. .10444)
                     /gene="ECU03_0090"
                     /function="unknown"
                     /note="03_0090, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26156.1"
                     /db_xref="GI:19168637"
                     /translation="MGECPFQTMRMWLVSAIALVDILCTSKIEELVENMGKKIEEALP
                     HEFDSAEKEALKKHLKETAGLGTRLMVPCIFHEDRVVASPTTRYQDIDEEEKGYVEKV
                     IALLPRLLWRSVAYIYVFGNDSWVVNLMEEVFETAPFKKSDAVALYKRARGRLGIRLI
                     DLVNRTFRHNIGMLNRFGQRLAQEAEAKIQEISSSLSDEEKRKEEKMLQIIKEYGESL
                     CTKEKQEEIIRAQEIMCDACACIWERDSNRESFVMETYSRHLYLRMIGSSMDVEEPLL
                     SYIDHRGLIDAYEKYKSIDIVAELIKQVFTERGCISDESVNDAVCGVREREEAEKMRG
                     EEERRKKEEESLRNTLELLRMEEKEKSKSRGKKKKGGKKGSGEVTAKMEEEKKDSEEV
                     EESAEAEVSLEEMAVGGARSKERSSKKKSRSKGHRYKVHKRVLRWTKSAERIKAELDE
                     GSEEKWRNKSIEEIEEQKKVHDIIEVCVLLRSLDANRFFVSTNRYMKDGTERWKMVGV
                     GIFEEGGEKKVGKVEVGLYRDKGEGSVIYHLMFKAMDTEKAGKGARSSFGKGDDVEGL
                     EEGAGELSDMSGFEYPKGVRSEIVKGGDAFKIVYRNPKDTSEVLRSLTVLQKAEVL"
     gene            11401. .12966
                     /gene="ECU03_0100"
     CDS             11401. .12966
                     /gene="ECU03_0100"
                     /note="03_0100, similarity to HYPOTHETICAL PROTEIN
                     Y461_MYCPN, gene found by Glimmer"
                     /codon_start=1
                     /product="similarity to HYPOTHETICAL PROTEIN Y461_MYCPN"
                     /protein_id="CAD26157.1"
                     /db_xref="GI:19168638"
                     /db_xref="GOA:Q8SW77"
                     /translation="MKPAYLYLILLFLPLSVSREIPDMRDILYEDFHKEMQDPPTPLI
                     IDLMSTDLFRRLDRVAQLGLRNLLFTHANITRGAHCRGTRTMSSIYVNILNKNSCSES
                     CITPVQSHAIQVASLLHDIGHGPYSHLFEYTVTGMVHSNPEIYHDDEKKDSAKPWSHE

  Query Match             24.8%;  Score 32;  DB 3;  Length 194439;
  Best Local Similarity   58.3%;  Pred. No. 28;
  Matches   56;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||| |||  || |  ||   |    ||||   |||| |||| |  ||||  |  ||
Db     127681 TGACCTTGTGCGTGTCGACATACATGTTTGCATCGTTGACAACAATGTGCCCGTTCTTGA 127622

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGT 96
              |  || || ||| ||  | ||  || |  |   |||
Db     127621 TCATCAGGATCTCTTCTTGGTTCTGGTCGGCAATGT 127586


RESULT 8
AK119844
LOCUS       AK119844                3048 bp    mRNA    linear   PLN 29-OCT-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:002-178-D02, full
            insert sequence.
ACCESSION   AK119844
VERSION     AK119844.1  GI:37989467
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
  MEDLINE   22752273
   PUBMED   12869764
REFERENCE   2
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
```

```
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakahama,Y.,
            Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K.,
            Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M.,
            Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C.,
            Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y.,
            Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M.,
            Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K.,
            Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J.,
            Yokomizo,S. and Yoshimura,A.
  TITLE     Collection, mapping, and annotation of 28K full-length cDNA clones
            from japonica rice
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3048)
  AUTHORS   Kikuchi,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
            Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
            305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
            Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT     This clone is one of the 32K full-length cDNA clones from japonica
            rice.
            URL : http://cdna01.dna.affrc.go.jp/cDNA/
            NIAS Rice Full-Length cDNA Project Team:  Kikuchi,S., Satoh,K.,
            Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
            Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,  Kojima,K., Namiki,T.,
            Ohneda,E., Yahagi,W., Suzuki,K.,  Li,C., Ohtsuki,K., Shishiki,T.,
            Yamamoto,M. and Nakahama,Y.
            FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
            Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M.,
            Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M.,
            Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J.,  Oka,M.,
            Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y.,  Tsunoda,Y., Ueda,M.,
            Xie,Q., Yokomizo,S., Yoshimura,A.,  Matsubara,K. and Murakami,K.
            Genome Exploration Research Group in Riken Genomic Sciences  Center
            and Genome Science Laboratory in Riken: Adachi,J.,  Aizawa,K.,
            Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  Hanagaki,T.,
            Hara,A., Hashizume,W., Hayashida,K.,  Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Iida,J.,  Imamura,K., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I.,  Kanagawa,S., Katoh,H., Kawai,J.,
            Kishikawa-Hirozane,T.,  Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S.,  Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M.,  Osato,N.,
            Ota,Y., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M.,
            Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
            Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
            Yasunishi,A. and  Hayashizaki,Y.
FEATURES             Location/Qualifiers
     source          1..3048
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="002-178-D02"
ORIGIN

  Query Match             24.7%;  Score 31.8;  DB 8;  Length 3048;
  Best Local Similarity   61.4%;  Pred. No. 31;
  Matches   51;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              ||  || ||  |  |  |  |||   ||   |||| | | |||| |||       || ||
Db        506 AGTTATCACAGCATGGCTCTCTGCACACCCTGACCGCCAACGAGCTGGCTTCCAGCTGTT 565

Qy         76 TGTTGTGGTGCTGTGACGTGTTG 98
              |||||||||||||||  || |||
Db        566 TGTTGTGGTGCTGTGGTGTTTTG 588
```

```
RESULT 9
AC138542/c
LOCUS       AC138542               90417 bp    DNA     linear   HTG 20-MAR-2004
DEFINITION  Magnaporthe grisea chromosome 7 clone 12G19, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC138542
VERSION     AC138542.2  GI:45597487
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 90417)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     The sequence of Magnaporthe grisea chromosome 7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90417)
  AUTHORS   Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr and Brown,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) Plant Pathology - Fungal Genomics
            Laboratory, North carolina State University, 840 Main Campus Drive,
            Raleigh, NC 27606, USA
REFERENCE   3  (bases 1 to 90417)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2004) Center for Integrated Fungal Research,
            North Carolina State University, Plant Pathology - 840 Main Campus
            Dr, Raleigh, NC 27606, USA
COMMENT     On Mar 20, 2004 this sequence version replaced gi:27552536.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    90417: contig of 90417 bp in length.
FEATURES             Location/Qualifiers
     source          1..90417
                     /organism="Magnaporthe grisea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:148305"
                     /chromosome="7"
                     /clone="12G19"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 2;  Length 90417;
  Best Local Similarity   54.9%;  Pred. No. 44;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          5 CTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGT 64
              |||| | || |   |  | | |||| |  ||||    || | | || |||  |||  | |
Db      88659 CTTCCGGTCCTCAGGCCTCCATCGAACCATACTCCATACCAGGTCCTCGAAGGAGTCGTT 88600

Qy         65 CGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGG 117
              |  || | ||   | ||| | |||    |||   ||||  ||    | ||| |
Db      88599 CTTGGGCGTTCGTTCGTGCTTCTGGTTGGTGCAATCCATGCGCGGAAGTCCTG 88547


RESULT 10
AC138544/c
LOCUS       AC138544               92770 bp    DNA     linear   HTG 20-MAR-2004
DEFINITION  Magnaporthe grisea chromosome 7 clone 15G13, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC138544
VERSION     AC138544.2  GI:45597492
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
```

```
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 92770)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     The sequence of Magnaporthe grisea chromosome 7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92770)
  AUTHORS   Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr and Brown,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) Plant Pathology - Fungal Genomics
            Laboratory, North carolina State University, 840 Main Campus Drive,
            Raleigh, NC 27606, USA
REFERENCE   3  (bases 1 to 92770)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2004) Center for Integrated Fungal Research,
            North Carolina State University, Plant Pathology - 840 Main Campus
            Dr, Raleigh, NC 27606, USA
COMMENT     On Mar 20, 2004 this sequence version replaced gi:27552538.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1    92770: contig of 92770 bp in length.
FEATURES             Location/Qualifiers
     source          1..92770
                     /organism="Magnaporthe grisea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:148305"
                     /chromosome="7"
                     /clone="15G13"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 2;  Length 92770;
  Best Local Similarity   54.9%;  Pred. No. 44;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          5 CTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGT 64
              |||| | || |   |  | | |||| |  ||||    || | | || |||  |||  | |
Db        468 CTTCCGGTCCTCAGGCCTCCATCGAACCATACTCCATACCAGGTCCTCGAAGGAGTCGTT 409

Qy         65 CGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGG 117
              |  || | ||   | ||| | |||    |||   ||||  ||    | ||| |
Db        408 CTTGGGCGTTCGTTCGTGCTTCTGGTTGGTGCAATCCATGCGCGGAAGTCCTG 356


RESULT 11
AC144889
LOCUS       AC144889               128428 bp    DNA     linear   HTG 11-JUN-2003
DEFINITION  Bos taurus clone RP42-541E10, WORKING DRAFT SEQUENCE, 6 ordered
            pieces.
ACCESSION   AC144889
VERSION     AC144889.2  GI:31581601
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 128428)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
```

```
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128428)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 128428)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 11, 2003 this sequence version replaced gi:31044309.
            -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                Center project name: eey
                Center clone name: 541E10

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 127382 bases at least Q40
                Consensus quality: 127698 bases at least Q30
                Consensus quality: 127848 bases at least Q20
                Insert size: 130000; agarose-fp
                Insert size: 127928; sum-of-contigs
                Quality coverage: 10.65x in Q20 bases; agarose-fp
                Quality coverage: 10.82x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1    33969: contig of 33969 bp in length
            *     33970    34069: gap of unknown length
            *     34070    53360: contig of 19291 bp in length
            *     53361    53460: gap of unknown length
            *     53461    63539: contig of 10079 bp in length
            *     63540    63639: gap of unknown length
            *     63640   112600: contig of 48961 bp in length
            *    112601   112700: gap of unknown length
            *    112701   118744: contig of 6044 bp in length
            *    118745   118844: gap of unknown length
            *    118845   128428: contig of 9584 bp in length.
FEATURES             Location/Qualifiers
     source          1..128428
```

```
                    /organism="Bos taurus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9913"
                    /clone="RP42-541E10"
                    /clone_lib="RP42"
     misc_feature   1..48703
                    /note="clone overlaps with GenBank Accession Number
                    AC095020 clone RP42-243E1 (center project name cje)"
     misc_feature   1..33969
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:left"
     misc_feature   34070..53360
                    /note="assembly_fragment"
     misc_feature   53461..63539
                    /note="assembly_fragment"
     misc_feature   63640..112600
                    /note="assembly_fragment"
     misc_feature   72761..128428
                    /note="clone overlaps with GenBank Accession Number
                    AC093397 clone RP42-405D15 (center project name cja)"
     misc_feature   112701..118744
                    /note="assembly_fragment"
     misc_feature   118845..128428
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:right"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 2;  Length 128428;
  Best Local Similarity   53.7%;  Pred. No. 44;
  Matches   65;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          8 CTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGG 67
              || | |||     |  | |||  | | |    |||| |   || | |||  | |   |
Db     111483 CTCCCCGTTTGCCTCCCCTCGGCCCTCAGCTCCGACCAGAGCCTCCACCTGGCTCTCCCA 111542

Qy         68 GGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTAGTTCAG 127
              || ||| |||||||   | ||  | ||  ||||||||  | | |  |||      | |||
Db     111543 GGCCTTGTTGTTGTCTAGTTGCTAAGTCATGTCCAACTCTGTGACCCCGTGGACTTGCAG 111602

Qy        128 C 128
              |
Db     111603 C 111603


RESULT 12
AC093397
LOCUS       AC093397               207821 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Bos taurus clone RP42-405D15, WORKING DRAFT SEQUENCE, 10 ordered
            pieces.
ACCESSION   AC093397
VERSION     AC093397.2  GI:21465397
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 207821)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207821)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 207821)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 19, 2002 this sequence version replaced gi:15277346.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cja
                 Center clone name: 405D15

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 205824 bases at least Q40
                 Consensus quality: 206436 bases at least Q30
                 Consensus quality: 206755 bases at least Q20
                 Insert size: 161000; agarose-fp
                 Insert size: 206921; sum-of-contigs
                 Quality coverage: 11.74x in Q20 bases; agarose-fp
                 Quality coverage: 9.14x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1   22661: contig of 22661 bp in length
            *      22662   22761: gap of unknown length
            *      22762   45605: contig of 22844 bp in length
            *      45606   45705: gap of unknown length
            *      45706   64208: contig of 18503 bp in length
            *      64209   64308: gap of unknown length
            *      64309   70434: contig of 6126 bp in length
            *      70435   70534: gap of unknown length
            *      70535  123863: contig of 53329 bp in length
            *     123864  123963: gap of unknown length
            *     123964  133638: contig of 9675 bp in length
            *     133639  133738: gap of unknown length
            *     133739  151887: contig of 18149 bp in length
            *     151888  151987: gap of unknown length
            *     151988  161829: contig of 9842 bp in length
            *     161830  161929: gap of unknown length
            *     161930  181367: contig of 19438 bp in length
            *     181368  181467: gap of unknown length
            *     181468  207821: contig of 26354 bp in length.
FEATURES             Location/Qualifiers
     source          1..207821
```

```
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="RP42-405D15"
                     /clone_lib="RP42"
     misc_feature    1. .22661
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22762. .45605
                     /note="assembly_fragment"
     misc_feature    45706. .64208
                     /note="assembly_fragment"
     misc_feature    64309. .70434
                     /note="assembly_fragment"
     misc_feature    70535. .123863
                     /note="assembly_fragment"
     misc_feature    123964. .133638
                     /note="assembly_fragment"
     misc_feature    133739. .151887
                     /note="assembly_fragment"
     misc_feature    151988. .161829
                     /note="assembly_fragment"
     misc_feature    161930. .181367
                     /note="assembly_fragment"
     misc_feature    175822. .207821
                     /note="clone overlaps with GenBank Accession Number
                     AC120515 clone RP42-98H15 (center project name dje)"
     misc_feature    181468. .207821
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match            24.3%;  Score 31.4;  DB 2;  Length 207821;
  Best Local Similarity   53.7%;  Pred. No. 44;
  Matches   65;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          8 CTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGG 67
              || | |||     |  | |||  | | |    |||| |   || | |||  | |   |
Db      38349 CTCCCCGTTTGCCTCCCCTCGGCCCTCAGCTCCGACCAGAGCCTCCACCTGGCTCTCCCA 38408

Qy         68 GGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTAGTTCAG 127
              || ||| |||||||   | ||  | ||  ||||||||  | | |  |||      | |||
Db      38409 GGCCTTGTTGTTGTCTAGTTGCTAAGTCATGTCCAACTCTGTGACCCCGTGGACTTGCAG 38468

Qy        128 C 128
              |
Db      38469 C 38469


RESULT 13
AF105113
LOCUS       AF105113                5832 bp    DNA     linear   BCT 09-SEP-1999
DEFINITION  Streptococcus pneumoniae type 19A putative oligosaccharide repeat
            unit transporter (cps19AJ) gene, partial cds; UDP-N-acetyl
            glucosamine-2-epimerase (cps19AK), glucose-1-phosphate thymidylyl
            transferase (cps19AL), dTDP-4-keto-6-deoxyglucose-3,5-epimerase
            (cps19AM), dTDP-glucose-4,6-dehydratase (cps19AN), and
            dTDP-L-rhamnose synthase (cps19AO) genes, complete cds; and AliA
            (aliA) gene, partial cds.
ACCESSION   AF105113
VERSION     AF105113.1  GI:4406246
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 5832)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Comparative genetics of capsular polysaccharide biosynthesis in
            Streptococcus pneumoniae types belonging to serogroup 19
  JOURNAL   J. Bacteriol. 181 (17), 5355-5364 (1999)
```

```
  MEDLINE   99395043
   PUBMED   10464207
REFERENCE   2  (bases 1 to 5832)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1998) Molecular Microbiology Unit, Women's and
            Children's Hospital, 72 King William Rd, Adelaide, SA 5006,
            Australia
FEATURES             Location/Qualifiers
     source          1. .5832
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="19A2"
                     /serotype="19A"
                     /db_xref="taxon:1313"
                     /note="serogroup: 19"
     gene            <1. .818
                     /gene="cps19AJ"
     CDS             <1. .818
                     /gene="cps19AJ"
                     /note="Cps19AJ"
                     /codon_start=3
                     /transl_table=11
                     /product="putative oligosaccharide repeat unit
                     transporter"
                     /protein_id="AAD19911.1"
                     /db_xref="GI:4406247"
                     /translation="SKFIFNIKESFDNIRVGINLMLSNIASSLIIGIVRLGIQWNWNI
                     ETFGKVSLTLSVSNLLMTFINAIGLVVFPLLRRTKAENLPKIYSNLRNVLMLIMFAIL
                     LIYYPLKIVLDLWLPAYQDALIFMTLIFPMSVYEGKMALVINTYLKALRMERDILRVN
                     ALVMLISMGVTLVTTYLLNSLELTVVSIVVLLALRSIIAELILSKKLDVSVKKDIVLE
                     FLLTLVFISSSWYLPIGLAVIVYTIAYGLYLYLKHEDIKTYLAYFKASKKTSN"
     gene            883. .1971
                     /gene="cps19AK"
     CDS             883. .1971
                     /gene="cps19AK"
                     /note="Cps19AK"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-N-acetyl glucosamine-2-epimerase"
                     /protein_id="AAD19912.1"
                     /db_xref="GI:4406248"
                     /translation="MKKIMLVFGTRPEAIKMCPLVNELKKHEDMETIVCVTGQHKEMV
                     SPVLDLFGVVPDYDLEIMKANQTLFSITTSILEKIKPVLEKEQPDIVLIHGDTTTTYA
                     AALAAFYLGIKVGHVEAGLRTYNLQSPFPEEFNRQSTSIIATYHFAPTELAKENLLKE
                     GRENVYVTGNTVIDALTTTVQEDYTHTHLDLNANNRLILLTAHRRENLGEPMRHMFRA
                     VKRVLNEYEDVKVIYPIHKNPLVRETAAEIFGDTERIQIIEPLDVLDFHNFMNQSYMI
                     LTDSGGVQEEAPSLGKPVLVMRDTTERPEGVAAGTLKLVGTDEETIYQNFKLLLDDSG
                     EYKKMSQASNPYGNGDASQQIVQILRGI"
     gene            2134. .3003
                     /gene="cps19AL"
     CDS             2134. .3003
                     /gene="cps19AL"
                     /note="Cps19AL"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidylyl transferase"
                     /protein_id="AAD19913.1"
                     /db_xref="GI:4406249"
                     /translation="MKGIILAGGSGTRLYPLTRAASKQLMPVYDKPMIYYPLSTLMLA
                     GIKDILIISTPQDLPRFKDLLLDGSEFGIKLSYAEQPSPDGLAQAFLIGEEFIGDDSV
                     ALILGDNIYHGPGLSTMLQKAAKKEKGATVFGYQVKDPERFGVVEFDTDMNAISIEEK
                     PEYPRSNYAVTGLYFYDNDVVEIAKQIKPSARGELEITDVNKAYLNRGDLSVELMGRG
                     FAWLDTGTHESLLEASQYIETVQRMQNVQVANLEEISYRMGYISREDVLELAQPLKKN
                     EYGRYLLRLIGEA"
     gene            3004. .3597
                     /gene="cps19AM"
     CDS             3004. .3597
                     /gene="cps19AM"
                     /note="Cps19AM"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-4-keto-6-deoxyglucose-3,5-epimerase"
```

```
                     /protein_id="AAD19914.1"
                     /db_xref="GI:4406250"
                     /translation="MSDNFFGKTLVVRKIDAIPGLLEFDIPVHGDNRGWFKENFQKEK
                     MLPLGFPESFFAAGKLQNNVSFSRKNVLRGLHAEPWDKYISVADDGKVLGSWVDLREG
                     ETFGNVYQTEIDASKGIFVPRGVANGFQVLSDTVSYSYLVNDYWALELKPKYAFVNYA
                     DPALGIEWENLPEAEVSEADKHHPLLRDVKPLTKDEL"
     gene            3609. .4655
                     /gene="cps19AN"
     CDS             3609. .4655
                     /gene="cps19AN"
                     /note="Cps19AN"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-glucose-4,6-dehydratase"
                     /protein_id="AAD19915.1"
                     /db_xref="GI:4406251"
                     /translation="MTEYKKIIVTGGAGFIGSNFVHYVYNNFPDVHVTVLDKLTYAGN
                     RANIEEILGDRVELVVGDIADAALVDKLAAEADAIVHYAAESHNDNSLNDPSPFIHTN
                     FIGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPGHGEGLGEKFTAETKYNPSSP
                     YSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILSGIKPKLY
                     GEGKNVRDWIHTNDHSSGVWTILTKGQIGETYLIGADGEKNNKEVLELILKEMGQPAD
                     AYDHVTDRAGHDLRYAIDASKLRDELGWKPEFTNFEAGLKETIKWYTDNQDWWKSEKE
                     AVEANYAKTQQVIK"
     stem_loop       4659. .4704
     gene            complement(4717. .5571)
                     /gene="cps19AO"
     CDS             complement(4717. .5571)
                     /gene="cps19AO"
                     /note="Cps19AO"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-L-rhamnose synthase"
                     /protein_id="AAD19916.1"
                     /db_xref="GI:4406252"
                     /translation="MILITGANGQLGTELRYLLNERNVDYVAVDVTEMDITNSEMVEK
                     VFAEVKPTLVYHCAAYTAVDAAEDEGKELDFAINVTGTENVAKGSEKYGAILVYISTD
                     YVFDGKKPVGQEWEVDDIPDPQTEYGRTKRMGEELVENLTSQHYIIRTAWVFGNYGRN
                     FVFTMQSLAETHKTLTVVSDQYGRPTWTRTLAEFMTYVTENQKEYGYYHLSNDAAEDT
                     TWYDFASEILKESDVELLPVDSSKFPAKAKRPLNSTMSLAKAKATGFVIPTWQDALKE
                     FYKQEVRK"
     gene            5817. .>5832
                     /gene="aliA"
     CDS             5817. .>5832
                     /gene="aliA"
                     /codon_start=1
                     /transl_table=11
                     /product="AliA"
                     /protein_id="AAD19917.1"
                     /db_xref="GI:4406253"
                     /translation="MKSSK"
ORIGIN

  Query Match             24.0%;  Score 31;  DB 1;  Length 5832;
  Best Local Similarity   59.8%;  Pred. No. 57;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||||  ||  |  |  ||| | || || | |  |||| ||||| | || || |
Db       2722 ATTACAGACGTTAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 2781

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db       2782 CGTGGTTTTGCCTGGTTGGATACGGGA 2808


RESULT 14
AF094575
LOCUS       AF094575               18754 bp    DNA     linear   BCT 09-SEP-1999
DEFINITION  Streptococcus pneumoniae serotype 19A DexB (dexB) gene, partial
            sequence; capsular polysaccharide biosynthesis operon, complete
            sequence; and oligopeptide permease AliA (aliA) gene, partial cds.
ACCESSION   AF094575
VERSION     AF094575.1  GI:3907597
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 18754)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Analysis of the 5' portion of the type 19A capsule locus identifies
            two classes of cpsC, cpsD, and cpsE genes in Streptococcus
            pneumoniae
  JOURNAL   J. Bacteriol. 181 (11), 3599-3605 (1999)
  MEDLINE   99287847
   PUBMED   10348877
REFERENCE   2  (bases 1 to 18754)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Comparative genetics of capsular polysaccharide biosynthesis in
            Streptococcus pneumoniae types belonging to serogroup 19
  JOURNAL   J. Bacteriol. 181 (17), 5355-5364 (1999)
  MEDLINE   99395043
   PUBMED   10464207
REFERENCE   3  (bases 1 to 18754)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1998) Microbial Microbiology Unit, Women's and
            Children's Hospital, King William Rd, North Adelaide, SA 5006,
            Australia
FEATURES             Location/Qualifiers
     source          1. .18754
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /serotype="19A"
                     /db_xref="taxon:1313"
     gene            <1. .9
                     /gene="dexB"
     CDS             <1. .9
                     /gene="dexB"
                     /codon_start=1
                     /transl_table=11
                     /product="DexB"
                     /protein_id="AAC78662.1"
                     /db_xref="GI:3907598"
                     /translation="VE"
     repeat_region   complement(161. .1647)
                     /note="insertion sequence; 3' truncated"
                     /rpt_family="IS1202"
                     /rpt_type=dispersed
     gene            1733. .16682
                     /gene="capsular polysaccharide biosynthesis operon"
     gene            1733. .3178
                     /gene="cps19aA"
     CDS             1733. .3178
                     /gene="cps19aA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein Cps19aA"
                     /protein_id="AAC78663.1"
                     /db_xref="GI:3907599"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFTYNILAF
                     RYLNLVVPALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSEIENVTQLMSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEHPDYASKIKKIYTKGFTKKVEAP
                     KTSKDQSFNIYVSGIDTYGPISSVSRSYVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKLIDLLGGVDVYNDQ
                     DFTSLHGKFHFPVGNVHLDSEQALGFVRERYSLADGDHDRGRNQQKVIAAILQKLTSS
                     EALKNYSRIIDSLQDSIQTNMPLETMINLVNAQLESGGTYKVNSQDLKGMGRTDLPSY
                     AMPDSNLYMMEINDSSLASVKTAIQDVLEGR"
     gene            3180. .3911
                     /gene="cps19aB"
     CDS             3180. .3911
                     /gene="cps19aB"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aB"
```

```
                     /protein_id="AAC78664.1"
                     /db_xref="GI:3907600"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVASDLVIAYGAEIYYTSDVLDKLEKNRIPTLNNS
                     RYALIEFSMNTPYRDIHSALSKILMLGITPVIAHIERYDALENNEKRVRELINMGCYT
                     QVNSSHVLKSKLFGERYKFMKKRAQYFLERDLVHIIASDMHNVDSRPPHMAEAYDLVS
                     QKYGEAKAQELFIDNPRKIVMDQLI"
     gene            3917. .4612
                     /gene="cps19aC"
     CDS             3917. .4612
                     /gene="cps19aC"
                     /function="chain length regulation and export"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aC"
                     /protein_id="AAC78665.1"
                     /db_xref="GI:3907601"
                     /translation="MMKEQNTIEIDVFQLLKTLWKHKLIILLVALVTGAGAFAYSTFI
                     VKPEYTSTTRIYVVNRNQGDKPGLTNQDLQAGTYLVKDYREIILSQDVLEKVATNLKL
                     DMPAKTLTSKVQVTVPADTRIVSISVKDKQPEEASRIANSLREVAAEKIIAVTRVSDV
                     TTLEEARPAMTPSSPNVRRNTLVGFLGGAAVTVITVLLIELLDTRVKRPEDVEDVLQM
                     PLLGVVPDFDKMK"
     gene            4622. .5311
                     /gene="cps19aD"
     CDS             4622. .5311
                     /gene="cps19aD"
                     /function="chain length regulation and export"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aD"
                     /protein_id="AAC78666.1"
                     /db_xref="GI:3907602"
                     /translation="MPTLEISQAKLELVKKAEEYYNALCTNLQLSGDDLKVFSITSVK
                     AGEGKTTTSTNIAWAFARAGYKTLLIDADMRNSVMSGVFKSRERITGLTEFLSGTTDL
                     SQGLCDTNIENLFVIQAGSVSPNPTALLQSKNFTTMLETLRKYFDYIVVDTAPVGVVI
                     DAAIIMQKCDASILVTKAGETKRRDLQKAKEQLEQTGKSCLGVVLNKLDTSVEKYGSY
                     GSYGNYRKQKK"
     gene            5326. .6687
                     /gene="cps19aE"
     CDS             5326. .6687
                     /gene="cps19aE"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosyl-1-phosphate transferase Cps19aE"
                     /protein_id="AAC78667.1"
                     /db_xref="GI:3907603"
                     /translation="MSEKLAKPSVAIVQSFLVILLTYLLSAVRETEIVSTTAIVLYIL
                     HYFVFYISDYGRNFFKRGYLIELVQTLKYILFFALAISISNFFLEDRFSISRRGMIYF
                     LLLHVFLVYVLNPFIKWYWKRAYPNFKGSKKILLLTATSRVEKVLDRLIESDDVVGEL
                     VAVSVLDKPDFQHDDLKVVAEGEIVNFATHEVVDEVFINLPSEKYNIGELVSQFETMG
                     IDVTVNLNAFDWARNKQICEMAGLNVVTFSTTFYKTSHVIAKRIIDIVGSLVGLILCG
                     LVSIVLVPLIRKDGGSAFFAQTRIGKNGRHFTFYKFRSMCVDAEEKKRELMEQNTMQG
                     GMFKVDDDPRITKIGHFIRKTSLDELPQFYNVLKGDMSLVGTRPPTVDEYEHYTPEQK
                     RRLSFKPGITGLWQVSGRSEIKNFDEVVKLDVAYIDGWTIWKDIEILLKTVKVVLMKD
                     GAK"
     gene            6693. .7436
                     /gene="cps19aF"
     CDS             6693. .7436
                     /gene="cps19aF"
                     /codon_start=1
                     /transl_table=11
                     /product="putative N-acetyl-mannosamine transferase
                     Cps19aF"
                     /protein_id="AAC78668.1"
                     /db_xref="GI:3907604"
                     /translation="MNERIQILGITIDPLTMKETVDAVEQYVLKKHPLHLMGVNADKI
                     NQCYEDEKIKKIVNESGIINADGASVVLASKFLGTPVPERVAGIDLMKHLLELSNEKG
                     YSVYFFGAKEDVLQDMLKIFEERYPALNVVGYRNGYFSPEDEKHIQEDIKEKKPDFVF
                     VGITSPKKEYIIQSFMDNGINAVFMGVGGSFDVLSGHIKRAPLWMQNAHLEWLFRVAN
                     EPKRLFKRYFVGNVSFIRKVLRAKRGYKL"
     gene            7475. .8257
                     /gene="cps19aG"
     CDS             7475. .8257
                     /gene="cps19aG"
                     /function="putative protein involved in formation of
                     phospho-diester linkage"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aG"
                     /protein_id="AAC78669.1"
                     /db_xref="GI:3907605"
                     /translation="MDAIKEFQKICKENNIDFFLRGGSVLGAVKYDGFIPWDDDMDIA
                     VPREGYDKLPGIFKDRIIAGKYQVLAYQYCDTLHCYFPRLFLLEDERKRLGLPRNTNL
                     GLHLIDIIPLDGAPNHSFLRKLYFGKVYWYRFLASLGTTYVGDHVDMHSAKQKLIIGF
                     FKKLGFAKLFPQNYVYRRLDNLYKKYDWKKQKYAGTINASLFAKEVMPVEIWGEGVEK
                     PFEDAFFKVPTEYDRYLKRLYGENYLHEEPSDDEKKSHLGGQ"
     gene            8259. .9137
                     /gene="cps19aH"
     CDS             8259. .9137
                     /gene="cps19aH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative rhamnosyl transferase Cps19aH"
                     /protein_id="AAC78670.1"
                     /db_xref="GI:3907606"
                     /translation="MFCYIILHYKVLEETISCVKSIKEGNSNTKQIVIIDNFSNNGTG
                     EKLQELYESDSEIDVLINHENAGFARGNNVAYQFAKEKYKPDFMVIMNNDIEIETEEF
                     EKIVTDIYRKEKFHLLGPDIFSTTYQLHQNPKRLTHYTYEEVKALNEKFKKGSQVSLA
                     LKIKCWLKSSKVLRTAIYQNRRKKKSVDYRKQVENPILHGSFIVYSRDLIEKEEYAFN
                     PNTFFYYETEILDYEAELKGYKRIYTPKIKVLHHQNVATNQVYTNLVEKTLFSNKCNF
                     ESTSYFLKLMKENEGV"
     gene            9138. .10472
                     /gene="cps19aI"
     CDS             9138. .10472
                     /gene="cps19aI"
                     /codon_start=1
                     /transl_table=11
                     /product="polysaccharide polymerase Cps19aI"
                     /protein_id="AAC78671.1"
                     /db_xref="GI:3907607"
                     /translation="MTYLFLLCLTLFLLTFFYFFAFNQDLIAPPVVMSVMFLISSVFA
                     LINVQNWNIEYSGLAYLLIISGIIVFSMPLLALNSPSLNTKIKVTDRLIDIQFWKIAL
                     TIIVDLIILYLYRREIHNLALSHGYTGSNFQWFFRNATSYEGELTVRTSIRVLIRIID
                     VSAYIFGYTFINNFFIYSHKRSKDLLLLVPFLIFISKTLLSGGRLDIIKILIAYVVMA
                     YIQQKRKVGWDKVISHKYMRLGFVGLIAGIPTFYYSLFLSGRSTTRTVFESISTYLGG
                     SIQHFNQYIQNPIGVAEVFGDESFVAIMNILGNLGFVNYNSTVHLEFRQLGITMGNVY
                     TFFRRPWHDFGLVGMYIFSFVVGVFFAIFYLKLRKSRAGFKLDIHTIIYSYFFYWIFL
                     SSIEQYSFTTISLFTLVFIVLVYLMAFFYWNLDFHRGKLVIKLSDTSIKSEN"
     gene            10497. .11924
                     /gene="cps19aJ"
     CDS             10497. .11924

  Query Match             24.0%;  Score 31;  DB 1;  Length 18754;
  Best Local Similarity   59.8%;  Pred. No. 58;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||||  ||  |  |  ||| | || || | |  |||| ||||| | || || |
Db      13828 ATTACAGACGTTAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 13887

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db      13888 CGTGGTTTTGCCTGGTTGGATACGGGA 13914


RESULT 15
AC096255/c
LOCUS       AC096255              222330 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-89M22, WORKING DRAFT SEQUENCE.
ACCESSION   AC096255
VERSION     AC096255.6  GI:30522200
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

```
            Rattus.
REFERENCE   1  (bases 1 to 222330)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 222330)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 222330)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942397.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
```

```
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GEUA
               Center clone name: CH230-89M22
            -------------- Summary Statistics
               Assembly program: Atlas 3.0;
               Consensus quality: 210552 bases at least Q40
               Consensus quality: 212815 bases at least Q30
               Consensus quality: 214416 bases at least Q20
               Estimated insert size: 220819; sum-of-contigs estimation
               Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   222330: contig of 222330 bp in length.
FEATURES             Location/Qualifiers
     source          1..222330
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-89M22"
     misc_feature    1..1784
                     /note="wgs_contig"
ORIGIN

  Query Match             23.9%;  Score 30.8;  DB 2;  Length 222330;
  Best Local Similarity   55.7%;  Pred. No. 69;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||| | || ||       |||  |  | |  ||    |   | || ||| || ||| |||
Db      81660 TGAGCATCCGCAATAGTGGATATCACCAAATATGCAAGGTAAAAACGACACCTACCCAGA 81601

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCG 106
              |  || | | |  | |||||||||  ||  || | |||  | | ||
Db      81600 TAATCTGAGGCACTGTGTTGTGGTTATGCAACATTTTGGACCATCG 81555

Search completed: August  6, 2005, 00:35:10
Job time : 1023.74 secs
```

Whitman, B
10/782899
Seq ID 1

```
                       GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          August  5, 2005, 21:21:24 ; Search time 1516.26 Seconds
                                            (without alignments)
                                            6135.749 Million cell updates/sec

Title:           US-10-782-899-1_COPY_1_192
Perfect score:   192
Sequence:        1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result                Query
  No.      Score  Match Length DB  ID                              Description
  -------------------------------------------------------------------------------
     1      192  100.0    600  6  BD144660                        BD144660 Drug for
c    2    175.2   91.2  14216  1  AE014814                        AE014814 Bifidobac
c    3    175.2   91.2 156638  6  AX492788                        AX492788 Sequence
c    4    175.2   91.2 156646  6  AX553955                        AX553955 Sequence
     5    173.2   90.2    600  1  AB072446                        AB072446 Bifidobac
c    6       37   19.3 230625  2  AC096816                        AC096816 Rattus no
c    7       37   19.3 246991  2  AC122956                        AC122956 Rattus no
     8     36.4   19.0  67395  9  AL671518                        AL671518 Human DNA
     9     36.4   19.0 285493  2  AC101665                        AC101665 Mus muscu
    10     35.8   18.6 170350  9  AC140952                        AC140952 Pan trogl
    11     35.8   18.6 171239  2  AC092873                        AC092873 Pan trogl
    12     35.8   18.6 171391  2  AC098646                        AC098646 Pan trogl
c   13     35.8   18.6 188467  2  AC151799                        AC151799 Pan trogl
c   14       35   18.2 181622  2  AC149460                        AC149460 Papio anu
    15       35   18.2 231844  2  AC116068                        AC116068 Rattus no
c   16       35   18.2 234234  2  AC097194                        AC097194 Rattus no
    17     34.8   18.1 152430  9  AC087432                        AC087432 Homo sapi
    18     34.8   18.1 160410  2  AC020753                        AC020753 Homo sapi
    19     34.8   18.1 166162  9  AC018814                        AC018814 Homo sapi
```

```
c   20     34.8   18.1 181870  2  AC024216                        AC024216 Homo sapi
    21     34.8   18.1 188432  2  AC018497                        AC018497 Homo sapi
    22     34.4   17.9 148269  2  AC026486                        AC026486 Homo sapi
    23     34.4   17.9 162249  2  AC020682                        AC020682 Homo sapi
c   24     34.4   17.9 166973  9  AL451049                        AL451049 Human DNA
c   25     34.2   17.8 155711  3  VYIVD10                         AL360354 Plasmodiu
    26     34.2   17.8 183089 10  AL663115                         AL663115 Mouse DNA
    27     34.2   17.8 213720 10  AC141477                         AC141477 Mus muscu
c   28       34   17.7 166873  9  AC008554                        AC008554 Homo sapi
    29       34   17.7 179393  9  AC010636                        AC010636 Homo sapi
    30     33.6   17.5   4799 10  AF367244                         AF367244 Mus muscu
    31     33.6   17.5  17073  3  AC084160                        AC084160 Caenorhab
    32     33.6   17.5 179619  2  AC109259                        AC109259 Mus muscu
c   33     33.4   17.4 114784  9  HS1189B24                       AL030996 Human DNA
    34     33.2   17.3    630 11  G54937                           G54937 SHGC-100131
    35     33.2   17.3   4765 10  BC052387                         BC052387 Mus muscu
    36     33.2   17.3   5011 10  AF140786                         AF140786 Mus muscu
    37     33.2   17.3   5034 10  AF387811                         AF387811 Mus muscu
    38     33.2   17.3  74436  2  AC026625                        AC026625 Homo sapi
    39     33.2   17.3  87258  9  AF091512                        AF091512 Homo sapi
    40     33.2   17.3 110000  8  CR382122_05                     Continuation (6 of
    41     33.2   17.3 132910  2  AC014319                        AC014319 Drosophil
    42     33.2   17.3 160421  9  AC116546                        AC116546 Homo sapi
    43     33.2   17.3 168700  2  AC009115                        AC009115 Homo sapi
    44     33.2   17.3 169557  3  AC104287                        AC104287 Drosophil
    45     33.2   17.3 180354  9  AC108134                        AC108134 Homo sapi


                              ALIGNMENTS


RESULT 1
BD144660
LOCUS       BD144660                 600 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Drug for gene therapy using anaerobic bacteria.
ACCESSION   BD144660
VERSION     BD144660.1  GI:27850418
KEYWORDS    JP 2002097144-A/1.
SOURCE      Bifidobacterium longum
  ORGANISM  Bifidobacterium longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Fujimori,M., Taniguchi,S., Amano,J., Yazawa,K., Kano,Y.,
            Nakamura,T. and Sasaki,T.
  TITLE     Drug for gene therapy using anaerobic bacteria
  JOURNAL   Patent: JP 2002097144-A 1 02-APR-2002;
            JUN AMANO,MINORU FUJIMORI
COMMENT     OS   Bifidobacterium longum
            PN   JP 2002097144-A/1
            PD   02-APR-2002
            PF   21-SEP-2001 JP 2001290187
            PI   MINORU FUJIMORI,SHUNICHIRO TANIGUCHI,JUN
            AMANO,KAZUTORA YAZAWA,
            PI   YASUMASA KANO,TOSHIYUKI NAKAMURA,TAKAYUKI SASAKI PC
            A61K35/74,A61K31/396,A61K31/7068,A61K31/711,A61K38/00,A61K38/ PC
            43,A61K48/00,
            PC   A61P35/00,C12N1/21,C12N15/09//(C12N1/21,C12R1:01),C12N15/00,
            PC   A61K37/02,
            PC   A61K37/48
            CC   Drug for gene therapy using anaerobic bacteria FH   Key
                 Location/Qualifiers
            FT   CDS             (193)..(471).
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:216816"
ORIGIN

  Query Match             100.0%;  Score 192;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-51;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 2
AE014814/c
LOCUS       AE014814               14216 bp    DNA     linear   BCT 31-OCT-2002
DEFINITION  Bifidobacterium longum NCC2705 section 201 of 202 of the complete
            genome.
ACCESSION   AE014814 AE014295
VERSION     AE014814.1  GI:23327097
KEYWORDS    .
SOURCE      Bifidobacterium longum NCC2705
  ORGANISM  Bifidobacterium longum NCC2705
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1  (bases 1 to 14216)
  AUTHORS   Schell,M.A., Karmirantzou,M., Snel,B., Vilanova,D., Berger,B.,
            Pessi,G., Zwahlen,M.-C., Desiere,F., Bork,P., Delley,M.,
            Pridmore,D. and Arigoni,F.
  TITLE     The genome sequence of Bifidobacterium longum reflects its
            adaptation to the human gastrointestinal tract
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14422-14427 (2002)
   PUBMED   12381787
REFERENCE   2  (bases 1 to 14216)
  AUTHORS   Schell,M.A., Karmirantzou,M., Snel,B., Vilanova,D., Berger,B.,
            Pessi,G., Zwahlen,M.-C., Desiere,F., Bork,P., Delley,M.,
            Pridmore,D. and Arigoni,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) Bioscience, Nestle Research Center, P.O.
            Box 44, Lausanne 26 1000, Switzerland
FEATURES             Location/Qualifiers
     source          1. .14216
                     /organism="Bifidobacterium longum NCC2705"
                     /mol_type="genomic DNA"
                     /strain="NCC2705"
                     /db_xref="taxon:206672"
     gene            178. .1638
                     /gene="BL1797"
     CDS             178. .1638
                     /gene="BL1797"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical proteasome-associated protein"
                     /protein_id="AAN25580.1"
                     /db_xref="GI:23327098"
                     /translation="MPQLRDSGTRSLHATEPVPSAEMDGFCRIFGVETEYGVAVTGAE
                     RPVDAGQVAMTMFQPIVSRSRSTNTYLANGSRLYLDVGSHPEYATAEARDPREALAQD
                     LAGEHVMRNLALKAQRKLRESYGAHATIHVFKNNVDSAGHAFGCHENYLVRRFVPLET
                     IEHQLLPFLITRQLYTGAGRMTPDGFQITQRADFLDEAVSSATTRSRPMVNTRDEPHA
                     DPDSFRRLHVIIGDSNRSQWSTWMKLAVTHLVLCAIEDAFRHGVPSGFEQYAFADPAA
                     ANRTVSRFLDNPRAELTLESGESVSALGLQRRYYAAVKAFIETHGDALAGSLPATTID
                     TIMGEWSRVLDALERGAYDALADRVDWAAKKCLFDALKRRRPDVTFAQMEQLELDYHD
                     IANGRLYGSLVARNQMRELLTGDNVEYAVHNPPTDTRAALRGRFVDAALNVGAQFSAD
                     WTHLTLTAPERREAILLDPFEAEPTLGFEQLMEALN"
     gene            complement(1754. .2035)
                     /gene="hup"
                     /note="synonym: BL1798"
     CDS             complement(1754. .2035)
                     /gene="hup"
                     /note="HS1; Hup; COG family: bacterial nucleoid
                     DNA-binding protein; PFAM_ID: Bac_DNA_binding"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-binding protein Hu"
                     /protein_id="AAN25581.1"
                     /db_xref="GI:23327099"
                     /translation="MAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGL
                     KLTGLFSAERVKRAARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTE"
     gene            complement(2181. .4742)
                     /gene="BL1799"
     CDS             complement(2181. .4742)
                     /gene="BL1799"
                     /note="COG family: predicted integral membrane protein"
                     /codon_start=1
                     /transl_table=11
                     /product="narrowly conserved hypothetical membrane
                     protein"
                     /protein_id="AAN25582.1"
                     /db_xref="GI:23327100"
                     /translation="MTTEQDQSTTEDRPDKATSSTTDRPIASGDGRSASFAVSGPDSP
                     ATDDSNDSGPYIDDIAPRRTRDFGDLTRAGLSLLMAAVVMVFAVYLGGMTRGVESDAH
                     TAAQVINWLADFPSTVLTQLATIVIVIIVLAQLLLAREWLQAAVSALAMFAGYGMVWV
                     VSTAISSLNDFTLPMALVSAATSYGSGLLPDIYAGMASFLTAAGPRRTRSTVKWSWNI
                     LYAIAAVMVVLSWHSVTGMLVSMAAGRTVGMLIRFVVGTQNKGVWGQDLVAALKGIGL
                     DTASLIRHQEPGVGHGSLAVTLDDDLAEGSRIYDLDTTDNRRFIVSVIDAQTHTVGYL
                     KQLWDWVRFTSVSIRRDRSVRDAVQHHFAMLLGLHNIKLPAPAPYGIADTDESAILVL
                     DAHTIELPANLNTLTKADAVAYMRYLSVASRRGYTHRRITPDTLARLEDGTAVIAGWL
                     NGDNASGPANTALDKVQLLVLFAALIGVEPAVEAAREAWGDTTLTALAPFIQKVAVPS
                     PTRALESWDKQLLKGLRSRITALADEETVESAEPVTLARFSWRSMITMLLVIVAVVVV
                     FTQLKPEEIITALTNANPLMAVVTLAFGVCGWIGSSISLGSLMARHKRDNMGVFMSQV
                     AGGFATVSMPAGVGPSFVNLQFLRKSGYRNTQATAIMSAALVVYYAVYFSMLVIIGLF
                     TGRNMFSGAIPTNTLVIVLGVVVVVLSIAMMIPPLRHWVTRRLMPLAKTYINQLLDVL
                     SQPRQLTVSCLGALFQNATTGLAFWAALQAFGHSSNPIETTFVFLLAYALGSAVPTPG
                     GLGGVEAALTFAFVAVGVPQGVALSATLLHRVVFYWLRIPLGAAAMKWLDKHNLV"
     gene            complement(4874. .6319)
                     /gene="purB"
                     /note="synonym: BL1800"
     CDS             complement(4874. .6319)
                     /gene="purB"
                     /note="adenylosuccinase; PurB; COG family:
                     adenylosuccinate lyase; PFAM_ID: lyase_1"
                     /codon_start=1
                     /transl_table=11
                     /product="adenylosuccinate lyase"
                     /protein_id="AAN25583.1"
                     /db_xref="GI:23327101"
                     /translation="MEVMKLTDISPAIALTPLDGRYHKATAPLVEYMSEPALNRERMR
                     VEVEWMILLANGFEGNGNQPIVDGVKPFTAEEQAFLRAIPEDFGAEGIKQHAAHEAIT
                     HHDVKAVEYYIDDQIDKAPAELTNINDELKTLVHFACTSEDINNLSIARCVKNAMENV
                     WTPEFKRIIDHLAGKAEEFKDMPLLSLTHGQPATPTTLGKELAVYVYRLNRQLKHIEN
                     QEYLGKINGATGTFGAHLAACPDVDWIAVSREFVTNRMGLTWNPLTTQIESHDWQAEL
                     YSTVSHANRIMHNLCVDVWMYISRGVFAQVPVKGATGSSTMPHKVNPIRFENAEANFE
                     ISCSLLDTLSATLVESRWQRDLTDSTTQRNIGSALGYSLLALTNLMGGLESIHPNTHV
                     IERELDENWEVLGEPIQTAMRACELKGLPGMDKPYEKVKGLMRGRTINKEQVEQFIDQ
                     QSFDPETAARLKALTPATYTGVASQLVAFGR"
     gene            6443. .7084
                     /gene="BL1801"
     CDS             6443. .7084
                     /gene="BL1801"
                     /note="COG family: permeases of the major facilitator
                     superfamily"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical membrane protein with unknown
                     function"
                     /protein_id="AAN25584.1"
                     /db_xref="GI:23327102"
                     /translation="MPATTIHTPPKRPIAVSSLKWQPLDIAVGATLGAACGVIFWGFN
                     FAYMPISSVFGAILPGFASIVHAVWYFSGTLAVLILRKPGAALYVNLIGTLVETVIGS
                     QFDLGFVIISALLQGVFAEIPFALTRYRMFNLPISIASGALVALEYGTYLMLFRYVGV
                     SFVSPRGIVHMISEVTGGVLISGVMAWYLMLAIAKTGVLDRLASGRALIERRA"
```

```
     gene            complement(7220. .7585)
                     /gene="BL1802"
     CDS             complement(7220. .7585)
                     /gene="BL1802"
                     /note="COG family: DNA-directed RNA polymerase specialized
                     sigma subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25585.1"
                     /db_xref="GI:23327103"
                     /translation="MKTYQVIVRPGEKYWILEIPGIGFTQARTTAEIESMARDLITVM
                     TQDADFALTVETKLPQSVQEHLDEARRLRKEEAECRSNAARETRAAAQELHGMGLALQ
                     QIGDILGVSRQRAHQLVNA"
     gene            7610. .7855
                     /gene="BL1803"
     CDS             7610. .7855
                     /gene="BL1803"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25586.1"
                     /db_xref="GI:23327104"
                     /translation="MASASLQIMSARLLSISLCRGMGNHKSVELEYLMTPSLPDTFKR
                     DILAFCRLLHQFDEPSPTRHYIVNSFLDNSVKCGIYF"
     gene            complement(7852. .9468)
                     /gene="BL1804"
     CDS             complement(7852. .9468)
                     /gene="BL1804"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25587.1"
                     /db_xref="GI:23327105"
                     /translation="MGVVMSEHDYWQVESSVYGGVGYAPATLEEYAAIAKALDDEAAG
                     FAAIATAWESAALQLQSHRHSAPMCVTLQSGDPSAVVPGHVTAPYAALGNRCYDHATA
                     CQRLSDDLRGAADLLIRAHSLYSQAEMTARRMFTELLQAGTQAKPGYAAAGVAAVAAG
                     GFLAGWTIDGKPNPAWMSTFTYPFQEGVLSGAGGIIGGVPIGKSIAHTDEVNKAAGKI
                     ANFSGPAKDVVQGNHLDVREVQANADVVRASGSVAESMENLRRLAEERLGKIELNSGL
                     EYGTIAIQRYERSDGTNSWLVTIPGTDGQPDSPFGWAQNVELMSADQERRRKADSARM
                     VAEAMRQAGIGKDEPVALIGHSQGGIVAATLASDWAEEYTIEHVVTAGSPVANHPIPQ
                     RTWVTSVEIDDELVAALDGAANPVTDNWLTVQGHVSPAPAATPSTVHSDGSCTPGATP
                     ITGLTPYDAAPVAGSTNGRELSHWIKYHQAAYQNATDLGSPAVQRHEAHFQEVINGEL
                     KETRYYQGRMTQSATIAPSERTTEFSTFGG"
     gene            complement(9453. .9800)
                     /gene="BL1805"
     CDS             complement(9453. .9800)
                     /gene="BL1805"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25588.1"
                     /db_xref="GI:23327106"
                     /translation="MSLVLLAGYMNTYTSNPAPIDPTGGFAGNTVTPFDYGTCPCHAA
                     VQRSRDLVSVQRAATLAQSDQTAVAVNRSSNIVWTGKAADLFRSRLSRTAQYSASLAQ
                     DMETTHRLAWGSS"
     gene            complement(9933. .10583)
                     /gene="BL1806"
     CDS             complement(9933. .10583)
                     /gene="BL1806"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAN25589.1"
                     /db_xref="GI:23327107"
                     /translation="MSALGTYRPGLWRRIAVAGCRRGGSKPVRFDYVQSGHRIIERQY
                     QSGSGRVHGYHHAQGATAADRRAIRRGRANAHASVAASQGRHRHQCFDFFRADKDHPE
                     TGRGAEQRFRFRASPVGTAIREFFIQCQKGRRVDRRTEEAGRGADEGQSAIHRYPVEH
                     HPIRAESDAEQGYRGYQAVVSSRSQLFFFLTPSRVMALSRLFPGFGCRLWHLSNFR"
     gene            complement(10642. .11673)
                     /gene="BL1807"
     CDS             complement(10642. .11673)
                     /gene="BL1807"
```

```
                     /note="COG family: uncharacterized Bcr"

  Query Match             91.2%;  Score 175.2;  DB 1;  Length 14216;
  Best Local Similarity   97.4%;  Pred. No. 1.4e-45;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2229 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 2170

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2169 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 2110

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2109 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 2050

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db       2049 AAGAAGGATGCTTT 2036


RESULT 3
AX492788/c
LOCUS       AX492788               156638 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 1106 from Patent EP1227152.
ACCESSION   AX492788
VERSION     AX492788.1  GI:23338471
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   None.
  TITLE     Bacterial strain and genome of bifidobacterium
  JOURNAL   Patent: EP 1227152-A 1106 31-JUL-2002;
            Societe des Produits Nestle S.A. (CH)
FEATURES             Location/Qualifiers
     source          1. .156638
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="1106 seq in place of 1099 because seq 1 (2.256.638
                     bases splitted-in 7 more sequences.~seq 0001: from
                     0.000.001 to 0.349.980 length: 349980~seq 1100: from
                     0.300.001 to 0.649.980 length: 349980~seq 1101: from
                     0.600.001 to 0.949.980 length: 349980~seq 1102: from
                     0.900.001 to 1.249.980 length: 349980~seq 1103: from
                     1.200.001 to 1.549.980 length: 349980~seq 1104: from
                     1.500.001 to 1.849.980 length: 349980~seq 1105: from
                     1.800.001 to 2.149.980 length: 349980~seq 1106: from
                     2.100.001 to 2.256.638 length: 156638"
ORIGIN

  Query Match             91.2%;  Score 175.2;  DB 6;  Length 156638;
  Best Local Similarity   97.4%;  Pred. No. 1.6e-45;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142164 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 142105

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142104 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 142045

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142044 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 141985

Qy        179 AAGAAGGATGCTTT 192
```

```
                 ||||||||||||||
Db     141984 AAGAAGGATGCTTT 141971


RESULT 4
AX553955/c
LOCUS       AX553955               156646 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 9 from Patent WO02074798.
ACCESSION   AX553955
VERSION     AX553955.1  GI:25897908
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Arigoni,F., Delley,M., Mollet,B., Pridmore,R.D., Schell,M.A.,
            Pohl,T.G. and Zwahlen,M.C.
  TITLE     The genome of a bifidobacterium
  JOURNAL   Patent: WO 02074798-A 9 26-SEP-2002;
            SOCIETE DES PRODUITS NESTLE S.A. (CH)
FEATURES             Location/Qualifiers
     source          1..156646
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="seq 1: original length 2256646 splitted into~seq 1:
                     from 0.000.001 to 0.349.980~seq 3: from 0.300.001 to
                     0.649.980~seq 4: from 0.600.001 to 0.949.980~seq 5: from
                     0.900.001 to 1.249.980~seq 6: from 1.200.001 to
                     1.549.980~seq 7: from 1.500.001 to 1.849.980~seq 8: from
                     1.800.001 to 2.149.980~seq 9: from 2.100.001 to 2.256.646"
ORIGIN

  Query Match             91.2%;  Score 175.2;  DB 6;  Length 156646;
  Best Local Similarity   97.4%;  Pred. No. 1.6e-45;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142172 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 142113

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142112 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 142053

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142052 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 141993

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db     141992 AAGAAGGATGCTTT 141979


RESULT 5
AB072446
LOCUS       AB072446                  600 bp    DNA     linear   BCT 15-JAN-2003
DEFINITION  Bifidobacterium longum biovar Longum hup gene for histone-like HU
            protein HB1, complete cds.
ACCESSION   AB072446
VERSION     AB072446.1  GI:19918953
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Takeuchi,A., Matsumura,H. and Kano,Y.
  TITLE     Cloning and expression in Escherichia coli of a gene, hup, encoding
            the histone-like protein HU of Bifidobacterium longum
  JOURNAL   Biosci. Biotechnol. Biochem. 66 (3), 598-603 (2002)
  MEDLINE   21999816
   PUBMED   12005055
REFERENCE   2  (bases 1 to 600)
  AUTHORS   Kano,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) Yasunobu Kano, Kyoto Pharmaceutical
            University, Institute of Molecular and Cellular Biology for
            Pharmaceutical Sciences; 1, Shichono-cho, Misasagi, Yamashina-ku,
            Kyoto 6078412, Japan (E-mail:ykano@mb.kyoto-phu.ac.jp,
            Tel:81755954719, Fax:81755954799)
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1679"
     RBS             180..185
     gene            193..474
                     /gene="hup"
     CDS             193..474
                     /gene="hup"
                     /codon_start=1
                     /transl_table=11
                     /product="histone-like HU protein HB1"
                     /protein_id="BAB88690.1"
                     /db_xref="GI:19918954"
                     /translation="MAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGL
                     KLTGLFSAERVKRPARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTE"
ORIGIN

  Query Match             90.2%;  Score 173.2;  DB 1;  Length 600;
  Best Local Similarity   97.4%;  Pred. No. 5.6e-45;
  Matches  187;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          3 TGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTC 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTC 60

Qy         63 --AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
                |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 6
AC096816/c
LOCUS       AC096816               230625 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-171E18, *** SEQUENCING IN PROGRESS
            ***, 7 unordered pieces.
ACCESSION   AC096816
VERSION     AC096816.6  GI:30520628
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 230625)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
```

```
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230625)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 230625)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942762.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFJV
                Center clone name: CH230-171E18
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 197146 bases at least Q40
                Consensus quality: 202146 bases at least Q30
                Consensus quality: 205589 bases at least Q20
                Estimated insert size: 207699; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      6876: contig of 6876 bp in length
            *       6877      6976: gap of unknown length
            *       6977     35044: contig of 28068 bp in length
            *      35045     35144: gap of unknown length
            *      35145     60636: contig of 25492 bp in length
            *      60637     60736: gap of unknown length
            *      60737     72181: contig of 11445 bp in length
            *      72182     72281: gap of unknown length
            *      72282    227314: contig of 155033 bp in length
            *     227315    227414: gap of unknown length
            *     227415    228683: contig of 1269 bp in length
            *     228684    228783: gap of unknown length
            *     228784    230625: contig of 1842 bp in length.
FEATURES             Location/Qualifiers
     source          1..230625
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-171E18"
     misc_feature    224079..227314
                     /note="wgs_contig"
ORIGIN

  Query Match             19.3%;  Score 37;  DB 2;  Length 230625;
  Best Local Similarity   54.9%;  Pred. No. 1.2;
  Matches   73;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy         29 GACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAA 88
              || || |  | |  ||||| |    |||||    |   |   |    |||   |  | ||
Db     127530 GAAAAACCCATTTATGTCTAAGAGCTCTATGCAGATATCTAACTTCTAAAGGAACATAAA 127471

Qy         89 AACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATG 148
              ||  ||| | ||||  |  |||||  || |  ||| | ||||||  ||  | |   |  |
Db     127470 AATACTTTTCAAACATGTATTTTCTTAGTAGGATGAGGTAGTATACTTCTTTAACCCTAG 127411

Qy        149 GACTAAGCAAAAG 161
              ||||| ||| | |
Db     127410 GACTAGGCAGAGG 127398


RESULT 7
AC122956/c
LOCUS       AC122956               246991 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-17E8, WORKING DRAFT SEQUENCE.
ACCESSION   AC122956
VERSION     AC122956.4  GI:25137800
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 246991)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
```

```
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 246991)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 246991)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23666770.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GXPC
                Center clone name: CH230-17E8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225722 bases at least Q40
                Consensus quality: 228759 bases at least Q30
                Consensus quality: 230490 bases at least Q20
                Estimated insert size: 231948; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1   246991: contig of 246991 bp in length.
FEATURES             Location/Qualifiers
     source          1..246991
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-17E8"
     misc_feature    1..1165
                     /note="wgs_contig"
     misc_feature    complement(238413..239222)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH308764"
     misc_feature    244729..246991
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             19.3%;  Score 37;  DB 2;  Length 246991;
  Best Local Similarity   54.9%;  Pred. No. 1.2;
  Matches   73;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy         29 GACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAA 88
              || || |  | |  ||||| |    |||||    |   |   |    |||   |  | ||
Db       6776 GAAAAACCCATTTATGTCTAAGAGCTCTATGCAGATATCTAACTTCTAAAGGAACATAAA 6717

Qy         89 AACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATG 148
              ||  ||| | ||||  |  |||||  || |  ||| | ||||||  ||  | |   |  |
Db       6716 AATACTTTTCAAACATGTATTTTCTTAGTAGGATGAGGTAGTATACTTCTTTAACCCTAG 6657

Qy        149 GACTAAGCAAAAG 161
              ||||| ||| | |
Db       6656 GACTAGGCAGAGG 6644


RESULT 8
AL671518
LOCUS       AL671518                67395 bp    DNA     linear   PRI 16-FEB-2002
DEFINITION  Human DNA sequence from clone RP11-315C6 on chromosome 6, complete
            sequence.
ACCESSION   AL671518
VERSION     AL671518.3  GI:18855253
KEYWORDS    HTG.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67395)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 21, 2002 this sequence version replaced gi:18491488.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-315C6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-315C6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP1-71D21 is at 65396 in this sequence.
            The true right end of clone RP11-403I19 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .67395
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-315C6"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             19.0%;  Score 36.4;  DB 9;  Length 67395;
  Best Local Similarity   54.5%;  Pred. No. 1.8;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              || |||  | ||   || |  |||||| |||||| |       | |||| | | | |||
Db      15983 CAAAATAATTTCAATTTGGAATATTTTAAATACCCTTAATTGGAAAGATCTCATATCCCA 16042

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
              ||| | |    |||| | |||   | |  |  ||  ||| ||   ||  | ||| |
Db      16043 TATTATAAAGTGGTTCTAGCAACTATACACTATAAAATCCTTTGGGAAGATATGAATGTG 16102

Qy        155 GCAAAAGTGCTTGT 168
              || ||    | |||
Db      16103 GCTAAGCCTCATGT 16116


RESULT 9
AC101665
LOCUS       AC101665               285493 bp    DNA     linear   HTG 24-MAR-2004
DEFINITION  Mus musculus chromosome 17 clone RP23-55C11 map 17, *** SEQUENCING
            IN PROGRESS ***, 10 unordered pieces.
ACCESSION   AC101665
VERSION     AC101665.4  GI:45680671
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 285493)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 17, clone RP23-55C11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 285493)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 285493)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 24, 2004 this sequence version replaced gi:28475955.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
```

```
            Center project name: L17182
            Center clone name: 55_C_11
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   143428: contig of 143428 bp in length
            *   143429   143528: gap of 100 bp
            *   143529   155028: contig of 11500 bp in length
            *   155029   155128: gap of 100 bp
            *   155129   159909: contig of 4781 bp in length
            *   159910   160009: gap of 100 bp
            *   160010   170924: contig of 10915 bp in length
            *   170925   171024: gap of 100 bp
            *   171025   178641: contig of 7617 bp in length
            *   178642   178741: gap of 100 bp
            *   178742   204614: contig of 25873 bp in length
            *   204615   204714: gap of 100 bp
            *   204715   240582: contig of 35868 bp in length
            *   240583   240682: gap of 100 bp
            *   240683   246367: contig of 5685 bp in length
            *   246368   246467: gap of 100 bp
            *   246468   250304: contig of 3837 bp in length
            *   250305   250404: gap of 100 bp
            *   250405   285493: contig of 35089 bp in length.
FEATURES             Location/Qualifiers
     source          1..285493
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /map="17"
                     /clone="RP23-55C11"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

  Query Match             19.0%;  Score 36.4;  DB 2;  Length 285493;
  Best Local Similarity   54.5%;  Pred. No. 1.9;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         13 GGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCG 72
              |||  |   ||||||| ||| ||| |    || |||||     | ||  || || |||
Db     194920 GGCATTCTGGTGGCTTCACATGCAAATCACTGCCTGATGATAATGTTGCCAGTATCTTGC 194979

Qy         73 GGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTAT 132
                | | | | |||||| ||| | ||  |       |       ||| |   ||||  |||
Db     194980 ATGCATTGGGTGTGAATACCATAATGCAGTAGCCCTCCATTTTGAAGCCATCGCTGTTAT 195039

Qy        133 CATTGATGACAACA 146
              |  ||||  ||| |
Db     195040 CTCTGATTGCAAAA 195053


RESULT 10
AC140952
LOCUS       AC140952              170350 bp    DNA     linear   PRI 26-SEP-2003
DEFINITION  Pan troglodytes BAC clone RP43-22I12 from 7, complete sequence.
ACCESSION   AC140952
VERSION     AC140952.1  GI:28850254
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 170350)
  AUTHORS   Shah,N., Meyer,R. and Bielicki,L.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-22I12
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 170350)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Sequencing of Pan troglodytes
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 170350)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 170350)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT022I12
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..170350
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-22I12"
                     /clone_lib="RPCI-43"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 9;  Length 170350;
  Best Local Similarity   63.2%;  Pred. No. 2.9;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
              ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db     144944 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 145003

Qy         75 GAAATAGATGTGAAAACCCTTATAAAA 101
                ||    | ||||||| |||| ||||
```

```
Db     145004 ATAAGCATTTTGAAAACACTTAAAAAA 145030


RESULT 11
AC092873
LOCUS       AC092873              171239 bp    DNA     linear   HTG 12-JUN-2002
DEFINITION  Pan troglodytes clone RP43-22I12, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC092873
VERSION     AC092873.3  GI:21392490
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 171239)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171239)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 171239)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 12, 2002 this sequence version replaced gi:15799603.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ani
                 Center clone name: 022I12

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 170538 bases at least Q40
                 Consensus quality: 170688 bases at least Q30
                 Consensus quality: 170863 bases at least Q20
                 Insert size: 152000; agarose-fp
                 Insert size: 168000; pulse-field-gel
                 Insert size: 170939; sum-of-contigs
                 Quality coverage: 30.12x in Q20 bases; agarose-fp
                 Quality coverage: 27.25x in Q20 bases; pulse-field-gel
                 Quality coverage: 26.78x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1    82247: contig of 82247 bp in length
            *      82248    82347: gap of unknown length
            *      82348    83760: contig of 1413 bp in length
            *      83761    83860: gap of unknown length
            *      83861   168249: contig of 84389 bp in length
            *     168250   168349: gap of unknown length
            *     168350   171239: contig of 2890 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171239
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-22I12"
                     /clone_lib="RP43"
     misc_feature    1. .82247
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1. .62423
                     /note="clone overlaps with GenBank Accession Number
                     AC093139 clone RP43-22A1 (center project name anh)"
     misc_feature    61398. .171239
                     /note="clone overlaps with GenBank Accession Number
                     AC098646 clone RP43-44G6 (center project name anj)"
     misc_feature    82348. .83760
                     /note="assembly_fragment"
     misc_feature    83861. .168249
                     /note="assembly_fragment"
     misc_feature    168350. .171239
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 2;  Length 171239;
  Best Local Similarity   63.2%;  Pred. No. 2.9;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
               ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db      145797 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 145856

Qy          75 GAAATAGATGTGAAAACCCTTATAAAA 101
                 ||    | ||||||| |||| ||||
Db      145857 ATAAGCATTTTGAAAACACTTAAAAAA 145883


RESULT 12
AC098646
LOCUS       AC098646              171391 bp    DNA     linear   HTG 13-JUN-2002
DEFINITION  Pan troglodytes clone RP43-44G6, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC098646
VERSION     AC098646.2  GI:21397305
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 171391)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
```

```
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171391)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 171391)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:16506408.
            -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                Center project name: anj
                Center clone name: 044G06

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 170643 bases at least Q40
                Consensus quality: 170930 bases at least Q30
                Consensus quality: 171055 bases at least Q20
                Insert size: 150000; agarose-fp
                Insert size: 154000; pulse-field-gel
                Insert size: 171091; sum-of-contigs
                Quality coverage: 11.37x in Q20 bases; agarose-fp
                Quality coverage: 11.07x in Q20 bases; pulse-field-gel
                Quality coverage: 9.97x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    21570: contig of 21570 bp in length
            *    21571    21670: gap of unknown length
            *    21671    57729: contig of 36059 bp in length
            *    57730    57829: gap of unknown length
            *    57830   105820: contig of 47991 bp in length
            *   105821   105920: gap of unknown length .
            *   105921   171391: contig of 65471 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171391
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-44G6"
                     /clone_lib="RP43"
     misc_feature    1. .108807
                     /note="clone overlaps with GenBank Accession Number
                     AC092873 clone RP43-22I12 (center project name ani)"
     misc_feature    1. .21570
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1. .1027
                     /note="clone overlaps with GenBank Accession Number
                     AC093139 clone RP43-22A1 (center project name anh)"
     misc_feature    21671. .57729
                     /note="assembly_fragment"
     misc_feature    57830. .105820
                     /note="assembly_fragment"
     misc_feature    105921. .171391
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    138190. .171391
                     /note="clone overlaps with GenBank Accession Number
                     AC092871 clone RP43-120E8 (center project name ank)"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 2;  Length 171391;
  Best Local Similarity   63.2%;  Pred. No. 2.9;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
              ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db      83381 CCAGGCACTGGGTACATAAGTATCATATTTTTAATTTCTCTATGTTTAAAACATTTTAT 83440

Qy         75 GAAATAGATGTGAAAACCCTTATAAAA 101
                ||    | ||||||| |||| ||||
Db      83441 ATAAGCATTTGAAAACACTTAAAAAAA 83467


RESULT 13
AC151799/c
LOCUS       AC151799               188467 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Pan troglodytes chromosome 7 clone RP43-44D10, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC151799
VERSION     AC151799.1  GI:53386473
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 188467)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Pan troglodytes clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188467)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: C_PT044D10
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
```

```
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 186628 bases at least Q40
            Consensus quality: 187038 bases at least Q30
            Consensus quality: 187364 bases at least Q20
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1231: contig of 1231 bp in length
            *     1232      1331: gap of unknown length
            *     1332      3116: contig of 1785 bp in length
            *     3117      3216: gap of unknown length
            *     3217     24960: contig of 21744 bp in length
            *    24961     25060: gap of unknown length
            *    25061    102773: contig of 77713 bp in length
            *   102774    102873: gap of unknown length
            *   102874    187269: contig of 84396 bp in length
            *   187270    187369: gap of unknown length
            *   187370    188467: contig of 1098 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188467
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /clone="RP43-44D10"
     misc_feature    1. .1231
                     /note="assembly_name:Contig10"
     misc_feature    1332. .3116
                     /note="assembly_name:Contig13"
     misc_feature    3217. .24960
                     /note="assembly_name:Contig20
                     clone_end:T7
                     vector_side:left"
     misc_feature    25061. .102773
                     /note="assembly_name:Contig21
                     clone_end:SP6
                     vector_side:left"
     misc_feature    102874. .187269
                     /note="assembly_name:Contig22"
     misc_feature    187370. .188467
                     /note="assembly_name:Contig4"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 2;  Length 188467;
  Best Local Similarity   63.2%;  Pred. No. 2.9;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
              ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db     125311 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 125252

Qy         75 GAAATAGATGTGAAAACCCTTATAAAA 101
                ||    | ||||||| |||| ||||
Db     125251 ATAAGCATTTTGAAAACACTTAAAAAA 125225


RESULT 14
AC149460/c
LOCUS       AC149460               181622 bp    DNA     linear   HTG 08-JUL-2004
DEFINITION  Papio anubis clone RP41-441E1, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC149460
VERSION     AC149460.2  GI:50057884
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
```

```
SOURCE      Papio anubis (olive baboon)
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 181622)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
            Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P.,
            Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Jones,C.,
            Karlins,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
            Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
            Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C.,
            Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N.,
            Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D.,
            Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181622)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 181622)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 8, 2004 this sequence version replaced gi:48057351.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: hbl
                 Center clone name: 441E01

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 180653 bases at least Q40
                 Consensus quality: 180918 bases at least Q30
                 Consensus quality: 181167 bases at least Q20
                 Insert size: 180000; agarose-fp
                 Insert size: 181322; sum-of-contigs
                 Quality coverage: 10.99x in Q20 bases; agarose-fp
                 Quality coverage: 10.91x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     2622: contig of 2622 bp in length
```

```
            *     2623     2722: gap of unknown length
            *     2723   132243: contig of 129521 bp in length
            *   132244   132343: gap of unknown length
            *   132344   162447: contig of 30104 bp in length
            *   162448   162547: gap of unknown length
            *   162548   181622: contig of 19075 bp in length.
FEATURES             Location/Qualifiers
     source          1..181622
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-441E1"
                     /clone_lib="RP41"
                     /note="BAC resource: http://bacpac.chori.org/"
     misc_feature    1..2622
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2723..132243
                     /note="assembly_fragment"
     misc_feature    132344..162447
                     /note="assembly_fragment"
     misc_feature    132495..181622
                     /note="clone overlaps with GenBank Accession Number
                     AC150199 clone RP41-189I7 (center project name hbk)"
     misc_feature    162548..181622
                     /note="assembly_fragment
                     missing T7 clone end on 3' end of insert"
ORIGIN

  Query Match             18.2%;  Score 35;  DB 2;  Length 181622;
  Best Local Similarity   53.2%;  Pred. No. 5.3;
  Matches   74;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         42 TTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAA 101
              || |||  || | ||  | | |  | || |  | ||||  || ||   | |   |||||
Db      73060 TTTTCTACTTTGCCTCCTCTTACAAGCTACAAGCAAATGAATCTGCTTAGCAGAATAAAG 73001

Qy        102 CGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
              |     |||  ||| ||| ||    ||  || ||| |     | ||   ||||||   |
Db      73000 CCTATTTTTCAGCATAAATATTGCTTACCATGATTAAAATAGAAATCAGCTAAGCTTGAT 72941

Qy        162 TGCTTGTCCCCTGACCCAA 180
              | |||  |   ||  ||||
Db      72940 TTCTTTACTGGTGGACCAA 72922


RESULT 15
AC116068
LOCUS       AC116068              231844 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-66D24, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC116068
VERSION     AC116068.5  GI:30579201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231844)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

```
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231844)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231844)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24818872.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMWC
                Center clone name: CH230-66D24
            -------------- Summary Statistics
```

```
                Assembly program: Atlas 3.0;
                Consensus quality: 226512 bases at least Q40
                Consensus quality: 227588 bases at least Q30
                Consensus quality: 228638 bases at least Q20
                Estimated insert size: 242654; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   229543: contig of 229543 bp in length
            *   229544   229643: gap of unknown length
            *   229644   230670: contig of 1027 bp in length
            *   230671   230770: gap of unknown length
            *   230771   231844: contig of 1074 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231844
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-66D24"
     misc_feature    1. .2343
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(2941. .3750)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH264906"
     misc_feature    182235. .225520
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH264909"
     misc_feature    227512. .229543
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             18.2%;  Score 35;  DB 2;  Length 231844;
  Best Local Similarity   55.3%;  Pred. No. 5.4;
  Matches   68;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         14 GCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGG 73
              ||| | |||||  |    ||||| |      | |  |   | | ||| ||| | ||||
Db       9334 GCCTTTAAGTGTTTAATAAAGCAAACCACACTGTCTTATTTGTCTTTACAACAGCTTCAT 9393

Qy         74 GGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATC 133
               |||||  | || | |  || || ||  || |||||||  || |   |||   | |
Db       9394 TGAAATGTAAGTCACATGCCATAAAATTCGGGGGTTTTTCCACACTTATGAATTCGGTGG 9453

Qy        134 ATT 136
              |||
Db       9454 ATT 9456

Search completed: August  6, 2005, 00:35:05
Job time : 1522.26 secs
```

```
                        GenCore version 5.1.6
               Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 21:17:03 ; Search time 306.841 Seconds
                                           (without alignments)
                                           3704.166 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:      8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_16Dec04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002as:*
                7:  geneseqn2002bs:*
                8:  geneseqn2003as:*
                9:  geneseqn2003bs:*
                10:  geneseqn2003cs:*
                11:  geneseqn2003ds:*
                12:  geneseqn2004as:*
                13:  geneseqn2004bs:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                   Description
--------------------------------------------------------------------------------
    1     192   100.0    600  6  ABK52324             Abk52324 DNA encod
c   2   175.2    91.2 156638  6  ABQ81850             Abq81850 Bifidobac
c   3    33.4    17.4  50000  3  AAA64140             Aaa64140 Nucleotid
    4    32.6    17.0    278  3  AAC27526             Aac27526 Human sec
c   5    32.4    16.9   2717  4  ABL21958             Abl21958 Drosophil
    6      32    16.7 110000  6  ABQ69245_15          Continuation (16 o
c   7      32    16.7 110000  6  ABQ67195_2           Continuation (3 of
c   8    31.2    16.2  18272  4  AAK69447             Aak69447 Human imm
c   9    31.2    16.2  18272  4  AAL06181             Aal06181 Human rep
c  10    31.2    16.2  18272  4  ABL98746             Abl98746 Human tes
   11    30.6    15.9    466 13   ACN50973             Acn50973 Cotton an
c  12    30.6    15.9    549 12   ADJ44041             Adj44041 Plant cDN
   13    30.4    15.8    583  5  ABV51855             Abv51855 Human pro
   14    30.4    15.8   1569  4  AAH52605             Aah52605 S. epider
   15    30.4    15.8   3204  4  AAH54121             Aah54121 S. epider
   16    30.2    15.7    515  4  AAK37564             Aak37564 Human bon
   17    30.2    15.7 110000 11   ACN44014_2           Continuation (3 of
c  18      30    15.6   1332 13   ADS47100             Ads47100 Bacterial
   19      30    15.6   7017  4  AAS32482             Aas32482 Human gen
   20      30    15.6  25519 11   ACN44652             Acn44652 Mouse gen
```

```
c  21    29.8    15.5 115829 13   ABD33448            Abd33448 Human can
   22    29.6    15.4   2000  8  ADA71889             Ada71889 Rice gene
   23    29.6    15.4 260027 11   ACN44046             Acn44046 Human gen
   24    29.4    15.3    487  9  ACH25462             Ach25462 Human adu
   25    29.4    15.3    768 10   ADH85446             Adh85446 Enterococ
c  26    29.4    15.3    867 13   ADS57789             Ads57789 Bacterial
c  27    29.4    15.3   2000 10   ACC60804             Acc60804 Gene sequ
c  28    29.4    15.3   2000 10   ADK62093             Adk62093 Disease t
c  29    29.4    15.3   2414  4  ABL22762             Abl22762 Drosophil
c  30    29.4    15.3   2537  4  ABL11638             Abl11638 Drosophil
c  31    29.4    15.3   2719  4  ABL06226             Abl06226 Drosophil
c  32    29.4    15.3   3612 13   ADT47758             Adt47758 Bacterial
   33    29.4    15.3   4088  4  ABL03916             Abl03916 Drosophil
c  34    29.2    15.2    836  4  AAK83773             Aak83773 Human imm
c  35    29.2    15.2    836  4  AAK83776             Aak83776 Human imm
   36    29.2    15.2   1758  6  ABK87389             Abk87389 Target mo
   37    29.2    15.2   2190  8  ABZ79829             Abz79829 Human mat
   38    29.2    15.2   2250  8  ABZ79831             Abz79831 Human mat
   39    29.2    15.2   2253  8  ABZ79826             Abz79826 Human per
   40    29.2    15.2   2274  8  ABZ79830             Abz79830 Human mat
   41    29.2    15.2   2313  8  ABZ79828             Abz79828 Human TCG
   42    29.2    15.2   2337  8  ABZ79827             Abz79827 Human per
   43    29.2    15.2   2406 12   ADO42025             Ado42025 Human cel
   44    29.2    15.2   2508  8  ABZ79825             Abz79825 Human ost
   45    29.2    15.2   2655  5  AAS73346             Aas73346 DNA encod

                               ALIGNMENTS


RESULT 1
ABK52324
ID   ABK52324 standard; DNA; 600 BP.
XX
AC   ABK52324;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding cancer gene therapy associated protein.
XX
KW   Solid cancer; anaerobic bacteria; antitumour active protein; cancer;
KW   gene therapy; gene; ds.
XX
OS   Bifidobacterium longum.
XX
FH   Key             Location/Qualifiers
FT   CDS             193. .474
FT                   /*tag=  a
FT                   /product= "Cancer gene therapy protein"
XX
PN   JP2002097144-A.
XX
PD   02-APR-2002.
XX
PF   21-SEP-2001; 2001JP-00290187.
XX
PR   21-SEP-2000; 2000JP-00287688.
XX
PA   (AMAN/) AMANO A.
PA   (FUJI/) FUJIMORI M.
XX
DR   WPI; 2002-448201/48.
DR   P-PSDB; AAU96807.
XX
PT   Solid cancer therapy with anaerobic bacteria of Bifidobacterium sp. by
PT   tumor tissue specific delivery of a DNA encoding for an antitumor active
PT   protein or its precursor.
XX
PS   Claim 10; Page 16; 21pp; Japanese.
XX
CC   The invention describes a method of treating a solid cancer with
CC   anaerobic bacteria by site specific delivery of DNA encoding an
CC   antitumour active protein or its precursor. This sequence encodes a
```

```
CC   cancer gene therapy associated protein
XX
SQ   Sequence 600 BP; 132 A; 165 C; 161 G; 142 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 192;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-57;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 2
ABQ81850/c
ID   ABQ81850 standard; DNA; 156638 BP.
XX
AC   ABQ81850;
XX
DT   19-NOV-2002  (first entry)
XX
DE   Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1106.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS   Bifidobacterium longum.
OS   Synthetic.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-00102050.
XX
PR   30-JAN-2001; 2001EP-00102050.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT   a probe or primer for detecting and/or identifying Bifidobacterium longum
PT   in a biological sample.
XX
PS   Disclosure; SEQ ID NO 1106; 80pp; English.
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC   fusion protein, comprising a sequence selected from 1097 sequences given
CC   in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC   heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC   activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC   a probe) is useful for the detection and/or identification of
CC   Bifidobacterium longum in a biological sample. A carrier containing the
CC   lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be
```

```
CC   used for preventing and/or treating diarrhoea brought about by pathogenic
CC   bacteria and/or rotavirus. The carrier is a food composition selected
CC   from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC   products, ice-creams, fermented cereal based products, milk based
CC   powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tablets, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification. N.B. The sequence data for this patent is not represented
CC   in the printed specification but is based on sequence information
CC   supplied by the European Patent Office
XX
SQ   Sequence 156638 BP; 32098 A; 46491 C; 46415 G; 31634 T; 0 U; 0 Other;

  Query Match             91.2%;  Score 175.2;  DB 6;  Length 156638;
  Best Local Similarity   97.4%;  Pred. No. 3.1e-50;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142164 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 142105

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142104 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 142045

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142044 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 141985

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db     141984 AAGAAGGATGCTTT 141971


RESULT 3
AAA64140/c
ID   AAA64140 standard; DNA; 50000 BP.
XX
AC   AAA64140;
XX
DT   20-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of a beta-tubulin antigen.
XX
KW   Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;
KW   chronic ear disease; autoimmune disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200050593-A1.
XX
PD   31-AUG-2000.
XX
PF   25-FEB-2000; 2000WO-US004795.
XX
PR   25-FEB-1999;   99US-0121549P.
XX
PA   (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI   Yoo TJ;
XX
DR   WPI; 2000-558400/51.
XX
PT   New beta-tubulin antigen in the membranous structure of the inner ear,
PT   reactive with antibodies of patients with Meniere's disease, for
PT   diagnosing Meniere's disease and distinguishing this disease from other
PT   autoimmune ear diseases.
XX
PS   Claim 3; Page 74-97; 115pp; English.
```

```
XX
CC   The present sequence encodes a beta-tubulin antigen. The protein is an
CC   antigen of the membranous structure of the inner ear protein, and is
CC   reactive with antibodies from patients having Meniere's disease.
CC   Meniere's disease is a chronic ear disease with unknown etiology. Serum
CC   from patients suffering from this disease contain autoantibodies against
CC   a 30 kDa cochlear protein antigen. The disease is believed to be an
CC   autoimmune disease. The beta-tubulin antigen is useful as a target
CC   substance in diagnosing or detecting Meniere's disease and in
CC   distinguishing this disease from other autoimmune ear diseases
XX
SQ   Sequence 50000 BP; 17281 A; 9480 C; 8791 G; 14448 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 33.4;  DB 3;  Length 50000;
  Best Local Similarity   52.5%;  Pred. No. 2.8;
  Matches   73;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          42 TTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAA 101
               || |||  || | ||  | | |  | || |  | ||||  || ||   | |   || ||
Db       26376 TTTTCTACTTTGCCTCCTCTTACAAGCTACAAGCAAATGAATCTGCTCAGCAGAATGAAG 26317

Qy         102 CGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               |     |||  ||| ||| ||    |   || ||| |     | ||   |||||||  |
Db       26316 CCTATTTTTCAGCATAAATATTGCTTGCCATGATTAAAATAGAAATCAGCTAAGCATGAT 26257

Qy         162 TGCTTGTCCCCTGACCCAA 180
               | |||  |   ||  ||||
Db       26256 TTCTTTACTGGTGGACCAA 26238


RESULT 4
AAC27526
ID   AAC27526 standard; cDNA; 278 BP.
XX
AC   AAC27526;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 31601.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 31601; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 278 BP; 82 A; 44 C; 43 G; 108 T; 0 U; 1 Other;

  Query Match             17.0%;  Score 32.6;  DB 3;  Length 278;
  Best Local Similarity   51.0%;  Pred. No. 0.56;
  Matches   77;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          18 TGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAA 77
               | | |||| |||       |  |   || | |     || | |||||||  || |||||
Db          99 TCATGTGGATTGTTTTCTGTCTTTAGGTATTAAAATGCTGTATTCAATAATTTKGGGGAC 158

Qy          78 ATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTG 137
               ||| |||| |   | | |||| |  |  |  | ||   |  |||| |   |    || ||
Db         159 ATATATGTCATTTCACATATATACAGATGTGTATCAGGGTGACATTCCAAAAGTGCACTG 218

Qy         138 ATGACAACATGGACTAAGCAAAAGTGCTTGT 168
                |||      ||  || | |   ||  || |
Db         219 CTGAGTTAGAGGTATATGTATTTGTAATTTT 249


RESULT 5
ABL21958/c
ID   ABL21958 standard; DNA; 2717 BP.
XX
AC   ABL21958;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 17347.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 17347; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
```

```
SQ   Sequence 2717 BP; 835 A; 555 C; 503 G; 824 T; 0 U; 0 Other;

  Query Match             16.9%;  Score 32.4;  DB 4;  Length 2717;
  Best Local Similarity   56.6%;  Pred. No. 1.8;
  Matches   60;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         39 ATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATA 98
              ||  |  | ||| | |  |  | |||||| |      |  || || |  | |  |||| |
Db        369 ATAGTTACAGATACATAAAAATACAATACATAGCAAAAGGTACATATATAGAGACTTAAA 310

Qy         99 AAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              | ||||  |||||| | |||  || |  ||  ||  ||   |||||
Db        309 ATACGCTTGTTTTCACTGAATTATTCAATACGATTTTTAGAGACAA 264


RESULT 6
ABQ69245_15
Continuation (16 of 31) of ABQ69245 from base 1500001 (Listeria innocua DNA sequence #68
WP Sequence split into 31 fragments  LOCUS ABQ69245 Accession Abq69245
WP     Fragment Name      Begin      End
WP     ABQ69245_00            1      110000
WP     ABQ69245_01       100001      210000
WP     ABQ69245_02       200001      310000
WP     ABQ69245_03       300001      410000
WP     ABQ69245_04       400001      510000
WP     ABQ69245_05       500001      610000
WP     ABQ69245_06       600001      710000
WP     ABQ69245_07       700001      810000
WP     ABQ69245_08       800001      910000
WP     ABQ69245_09       900001     1010000
WP     ABQ69245_10      1000001     1110000
WP     ABQ69245_11      1100001     1210000
WP     ABQ69245_12      1200001     1310000
WP     ABQ69245_13      1300001     1410000
WP     ABQ69245_14      1400001     1510000
WP     ABQ69245_15      1500001     1610000
WP     ABQ69245_16      1600001     1710000
WP     ABQ69245_17      1700001     1810000
WP     ABQ69245_18      1800001     1910000
WP     ABQ69245_19      1900001     2010000
WP     ABQ69245_20      2000001     2110000
WP     ABQ69245_21      2100001     2210000
WP     ABQ69245_22      2200001     2310000
WP     ABQ69245_23      2300001     2410000
WP     ABQ69245_24      2400001     2510000
WP     ABQ69245_25      2500001     2610000
WP     ABQ69245_26      2600001     2710000
WP     ABQ69245_27      2700001     2810000
WP     ABQ69245_28      2800001     2910000
WP     ABQ69245_29      2900001     3010000
WP     ABQ69245_30      3000001     3011208

  Query Match             16.7%;  Score 32;  DB 6;  Length 110000;
  Best Local Similarity   50.7%;  Pred. No. 13;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db      29778 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 29837

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               || |  |   |  | || |||||     ||| | |   |||    |  || | | |
Db      29838 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 29897

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
              ||    ||    |   ||  | | ||  | ||
Db      29898 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 29929


RESULT 7
ABQ67195_2/c
Continuation (3 of 5) of ABQ67195 from base 200001 (Listeria innocua contig DNA sequence
WP Sequence split into 5 fragments  LOCUS ABQ67195 Accession Abq67195
WP     Fragment Name      Begin      End
WP     ABQ67195_0             1      110000
WP     ABQ67195_1        100001      210000
WP     ABQ67195_2        200001      310000
WP     ABQ67195_3        300001      410000
WP     ABQ67195_4        400001      495269

  Query Match             16.7%;  Score 32;  DB 6;  Length 110000;
  Best Local Similarity   50.7%;  Pred. No. 13;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db      29177 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 29118

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               || |  |   |  | || |||||     ||| | |   |||    |  || | | |
Db      29117 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 29058

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
              ||    ||    |   ||  | | ||  | ||
Db      29057 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 29026


RESULT 8
AAK69447/c
ID   AAK69447 standard; DNA; 18272 BP.
XX
AC   AAK69447;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24259.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
```

```
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
```

```
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 24259; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 18272 BP; 3601 A; 5274 C; 5400 G; 3997 T; 0 U; 0 Other;

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 11;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
```

```
Db      17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 9
AAL06181/c
ID   AAL06181 standard; DNA; 18272 BP.
XX
AC   AAL06181;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8869.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
```

```
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 8869; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 18272 BP; 3601 A; 5274 C; 5400 G; 3997 T; 0 U; 0 Other;

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 11;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 10
ABL98746/c
ID   ABL98746 standard; DNA; 18272 BP.
XX
AC   ABL98746;
XX
DT   21-JUN-2002  (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 3398.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001329.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
```

```
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483232/52.
XX
PT   Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT   for preventing, diagnosing and/or treating testicular cancer.
```

```
XX
PS   Disclosure; SEQ ID NO 3398; 766pp; English.
XX
CC   The present invention provides the protein and coding sequences of 973
CC   human testicular antigens, and fragments of their genomic sequences. The
CC   sequences can be used in the treatment of cardiovascular, urinary system,
CC   reproductive system, immune, respiratory, neurological and
CC   gastrointestinal disorders, infections, and particularly cancer,
CC   especially testicular cancers. The present sequence is a DNA encoding a
CC   protein fragment of the invention
XX
SQ   Sequence 18272 BP; 3601 A; 5274 C; 5400 G; 3997 T; 0 U; 0 Other;

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 11;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
               || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db       17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy         131 ATCATTGATGACAACATGGACTAA 154
               | ||   |||| |||| ||||| |
Db       17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 11
ACN50973
ID   ACN50973 standard; cDNA; 466 BP.
XX
AC   ACN50973;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Cotton androecium tissue EST Clone ID: LIB3828-002-Q1-N6-F6, SEQ:5754.
XX
KW   Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
KW   variety Nucotton33B; library LIB3828; molecular tag; molecular marker;
KW   genetic mapping; molecular mapping; seed germination; plant growth;
KW   plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS   Gossypium hirsutum.
XX
PN   US2004123340-A1.
XX
PD   24-JUN-2004.
XX
PF   12-DEC-2001; 2001US-00021323.
XX
PR   14-DEC-2000; 2000US-0255619P.
XX
PA   (DEIK/) DEIKMAN J.
PA   (FENG/) FENG P C C.
PA   (FINC/) FINCHER K L.
PA   (ZIEG/) ZIEGLER T E.
XX
PI   Deikman J,  Feng PCC,  Fincher KL,  Ziegler TE;
XX
DR   WPI; 2004-479808/45.
XX
PT   New isolated nucleic acid molecule that encodes a plant protein or its
PT   fragment, useful for isolating a variety of agronomically significant
PT   genes associated with plant growth, quality or yield, and as molecular
PT   tags to map genes.
XX
PS   Claim 1; SEQ ID NO 5754; 34pp; English.
XX
CC   The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC   ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC   from primed or non-primed seeds from variety DP50B, mature seeds from
CC   variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC   tissue, developing fibres, carpel walls and septa from variety
CC   Nucotton33B. The invention also relates to substantially purified
```

```
CC   proteins or their fragments encoded by nucleic acid molecules of the
CC   invention, and to transformed plants having a nucleic acid construct
CC   comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC   molecular tags to isolate genetic regions, to isolate genes, to map
CC   genes, to determine gene function and to determining whether genes are
CC   members of a particular gene family. The nucleic acid molecules may be
CC   used for isolating a variety of agronomically significant genes
CC   associated with plant growth, quality, yield, and could also serve as
CC   links in metabolic and catabolic pathways. The nucleic acid molecules are
CC   also useful for identifying genes important in initiating and maintaining
CC   seed germination or that may be used to mitigate stresses encountered
CC   during seed germination. The ESTs additionally enable the acquisition of
CC   promoters and cis-regulatory elements which will be useful to express
CC   agronomically significant genes in these tissues and/or other tissues,
CC   and also permits the acquisition of molecular markers useful in breeding
CC   schemes, genetic and molecular mapping, and in cloning of agronomically
CC   significant genes. The nucleic acid molecules are further useful for
CC   detecting the expression level or pattern of a protein or mRNA and for
CC   detecting the presence or quantity of a protein by tissue printing. The
CC   present sequence represents a specifically claimed EST isolated from a
CC   cotton variety Nucotton33B androecium tissue cDNA library (LIB3828). The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the US
CC   patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ   Sequence 466 BP; 138 A; 105 C; 81 G; 142 T; 0 U; 0 Other;

  Query Match             15.9%;  Score 30.6;  DB 13;  Length 466;
  Best Local Similarity   58.1%;  Pred. No. 3.6;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              | ||| ||   |   |  | || ||   | |||| | |   | | |  |  || ||||||
Db         27 ATATATATCATATGTCTTTCATTAACTTCTGGTTCTGGACTACATAATCCTTATTATCAT 86

Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
              |||| | | |  |||  || ||| ||  |||||
Db         87 TGATCATACCCGGGAGGAACCAAGAGCCCTTGT 119


RESULT 12
ADJ44041/c
ID   ADJ44041 standard; cDNA; 549 BP.
XX
AC   ADJ44041;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #5041.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/) BUDWORTH P.
PA   (MOUG/) MOUGHAMER T.
PA   (BRIG/) BRIGGS S P.
PA   (COOP/) COOPER B.
PA   (GLAZ/) GLAZEBROOK J.
PA   (GOFF/) GOFF S A.
PA   (KATA/) KATAGIRI F.
PA   (KREP/) KREPS J.
PA   (PROV/) PROVART N.
PA   (RICK/) RICKE D.
PA   (ZHUT/) ZHU T.
XX
PI   Budworth P,  Moughamer T,  Briggs SP,  Cooper B,  Glazebrook J;
PI   Goff SA,  Katagiri F,  Kreps J,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Example 13; SEQ ID NO 5041; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 549 BP; 136 A; 143 C; 88 G; 177 T; 0 U; 5 Other;

  Query Match             15.9%;  Score 30.6;  DB 12;  Length 549;
  Best Local Similarity   51.9%;  Pred. No. 3.9;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         48 GATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGG 107
              || || | |   ||||| ||    |  |||| || | |  ||   |||    ||   | |
Db        235 GAATCATTTGAGTTCAAGACTGCTGAAGAAAAAGTTATCGAAGGTCTTGACCAAGCAGTG 176

Qy        108 TTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTG 167
               |  ||  |||  | |      |  | ||| ||||||      || |  |    | ||||
Db        175 ATGACGATGAATAAGGGAGAGATTGCTTTGGTGACAATTCCACCTGAATATGCCTTCTTG 116

Qy        168 TCCCCTGACCCAA 180
              ||| ||||  |||
Db        115 TCCACTGAATCAA 103


RESULT 13
ABV51855
ID   ABV51855 standard; cDNA; 583 BP.
XX
AC   ABV51855;
XX
DT   17-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 51846.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 10067-10068; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 583 BP; 210 A; 79 C; 104 G; 188 T; 0 U; 2 Other;

  Query Match             15.8%;  Score 30.4;  DB 5;  Length 583;
  Best Local Similarity   52.9%;  Pred. No. 4.7;
  Matches   64;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              | || ||||   | |||     ||||||   || |   |||| |  |||    |  | ||
Db         18 AGTCTTCTAGNATTAATTAACGCGGGGATTGAGTTCATAAAAGCTATATTTCTCATGTGT 77

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
              ||  |||||||  | |   | ||||||   | | |    ||   ||   |||    || |
Db         78 TTGGGCAGAAATTTTCAAGAATATCATATTTCAGATTTAGGGAAAACTCAAAAGTTTTAT 137

Qy        169 C 169
              |
Db        138 C 138

RESULT 14
AAH52605
ID   AAH52605 standard; DNA; 1569 BP.
XX
AC   AAH52605;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis open reading frame nucleotide sequence SEQ ID NO:603.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW   endocarditis; ds.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809-A2.
```

```
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US030782.
XX
PR   09-NOV-1999;   99US-0164258P.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
DR   P-PSDB; AAG81755.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis.
XX
PS   Claim 8; Page 195; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC   and (II) can have antibacterial activity and therefore can be used in
CC   vaccination. The nucleic acids (I) may be used to produce the S.
CC   epidermidis polypeptides (II) via the production of vectors containing
CC   them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC   represent oligonucleotide sequences and primers which are used in the
CC   exemplification of the present invention. N.B. The present invention
CC   specifically claims all the polynucleotide sequences given in the
CC   sequence listing of the present specification, however the sequence
CC   listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC   in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC   for SEQ ID NO:4455 to 4464
XX
SQ   Sequence 1569 BP; 616 A; 186 C; 274 G; 493 T; 0 U; 0 Other;

  Query Match             15.8%;  Score 30.4;  DB 4;  Length 1569;
  Best Local Similarity   57.3%;  Pred. No. 7.2;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              ||  |  |||   ||| |   |  | ||   | || |   |   ||||||||||    ||
Db         87 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 146

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              |  ||| || | ||  | |||| | ||| |||| ||
Db        147 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 182

RESULT 15
AAH54121
ID   AAH54121 standard; DNA; 3204 BP.
XX
AC   AAH54121;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis genomic polynucleotide sequence SEQ ID NO:3485.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW   endocarditis; ds.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809-A2.
XX
PD   17-MAY-2001.
```

```
XX
PF   09-NOV-2000; 2000WO-US030782.
XX
PR   09-NOV-1999;   99US-0164258P.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis.
XX
PS   Claim 8; Page 1055-1056; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC   and (II) can have antibacterial activity and therefore can be used in
CC   vaccination. The nucleic acids (I) may be used to produce the S.
CC   epidermidis polypeptides (II) via the production of vectors containing
CC   them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC   represent oligonucleotide sequences and primers which are used in the
CC   exemplification of the present invention. N.B. The present invention
CC   specifically claims all the polynucleotide sequences given in the
CC   sequence listing of the present specification, however the sequence
CC   listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC   in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC   for SEQ ID NO:4455 to 4464
XX
SQ   Sequence 3204 BP; 1249 A; 411 C; 554 G; 990 T; 0 U; 0 Other;

  Query Match             15.8%;  Score 30.4;  DB 4;  Length 3204;
  Best Local Similarity   57.3%;  Pred. No. 9.8;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              ||  |  |||   ||| |   |  | ||   | || |   |   ||||||||||    ||
Db        542 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 601

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              |  ||| || | ||  | |||| | ||| |||| ||
Db        602 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 637

Search completed: August  5, 2005, 23:52:40
Job time : 310.841 secs
```

```
                     GenCore version 5.1.6
            Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 23:44:09 ; Search time 420.486 Seconds
                                           (without alignments)
                                           2959.924 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       7297361 seqs, 3241162794 residues

Total number of hits satisfying chosen parameters:      14594722

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
                18:  /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
                19:  /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
                20:  /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
                21:  /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
                22:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                23:  /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
                24:  /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
                25:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                26:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                      %
Result              Query
  No.     Score     Match Length DB  ID                     Description
---------------------------------------------------------------------------
    1       192    100.0     600  9  US-09-816-391A-1        Sequence 1, Appli
    2       192    100.0     600 21  US-10-782-899-1          Sequence 1, Appli
c   3     175.2     91.2 2256646  19  US-10-470-565-1          Sequence 1, Appli
c   4      34.6     18.0     425 20  US-10-425-115-60228      Sequence 60228, A
c   5      33.2     17.3     824 13  US-10-027-632-167810     Sequence 167810,
c   6      33.2     17.3     824 17  US-10-027-632-167810     Sequence 167810,
    7      32.4     16.9     484 13  US-10-027-632-6585       Sequence 6585, Ap
    8      32.4     16.9     484  17  US-10-027-632-6585         Sequence 6585, Ap
c   9        32     16.7  495269  17  US-10-398-221-8            Sequence 8, Appli
   10        32     16.7 3011208  17  US-10-398-221-2058          Sequence 2058, Ap
c  11      31.6     16.5     732  20  US-10-425-115-71957        Sequence 71957, A
c  12      31.2     16.2     360  19  US-10-674-124A-12097       Sequence 12097, A
c  13      31.2     16.2   18272  10  US-09-764-891-8869         Sequence 8869, Ap
   14      30.6     15.9     466  19  US-10-021-323-5754         Sequence 5754, Ap
c  15      30.6     15.9     549  17  US-10-260-238-5041         Sequence 5041, Ap
c  16      30.6     15.9    1099  13  US-10-027-632-10536        Sequence 10536, A
c  17      30.6     15.9    1099  17  US-10-027-632-10536        Sequence 10536, A
   18      30.4     15.8     583  20  US-10-357-930-51874        Sequence 51874, A
   19      30.2     15.7     515  9  US-09-864-761-16206        Sequence 16206, A
   20      30.2     15.7  370469  13  US-10-087-192-250          Sequence 250, App
c  21        30     15.6     201  20  US-10-719-993-41046        Sequence 41046, A
c  22        30     15.6     639  13  US-10-027-632-185956       Sequence 185956,
c  23        30     15.6     639  17  US-10-027-632-185956       Sequence 185956,
c  24        30     15.6    1332  17  US-10-369-493-25530        Sequence 25530, A
   25        30     15.6    7017  17  US-10-074-024-436          Sequence 436, App
   26        30     15.6   25519  13  US-10-087-192-1207         Sequence 1207, Ap
c  27        30     15.6  303172  20  US-10-719-993-6890         Sequence 6890, Ap
c  28      29.8     15.5     501  13  US-10-027-632-270844       Sequence 270844,
c  29      29.8     15.5     501  17  US-10-027-632-270844       Sequence 270844,
c  30      29.8     15.5   10303  17  US-10-264-213-105          Sequence 105, App
   31      29.8     15.5   10304  17  US-10-264-213-120          Sequence 120, App
c  32      29.8     15.5  115829  19  US-10-322-281-590          Sequence 590, App
c  33      29.6     15.4     453  13  US-10-027-632-266872       Sequence 266872,
c  34      29.6     15.4     453  17  US-10-027-632-266872       Sequence 266872,
   35      29.6     15.4    2510  18  US-10-425-114-24639        Sequence 24639, A
   36      29.6     15.4    2526  20  US-10-425-115-44214        Sequence 44214, A
   37      29.6     15.4  260027  13  US-10-087-192-298          Sequence 298, App
   38      29.4     15.3     487  10  US-09-918-995-12674        Sequence 12674, A
c  39      29.4     15.3     614  13  US-10-027-632-221811       Sequence 221811,
c  40      29.4     15.3     614  13  US-10-027-632-221812       Sequence 221812,
c  41      29.4     15.3     614  17  US-10-027-632-221811       Sequence 221811,
c  42      29.4     15.3     614  17  US-10-027-632-221812       Sequence 221812,
c  43      29.4     15.3     867  17  US-10-369-493-33463        Sequence 33463, A
c  44      29.4     15.3    3612  17  US-10-369-493-46196        Sequence 46196, A
   45      29.2     15.2     507  9  US-09-796-692-4795         Sequence 4795, Ap


                                   ALIGNMENTS


RESULT 1
US-09-816-391A-1
; Sequence 1, Application US/09816391A
; Patent No. US20020054865A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, Minoru
;  APPLICANT:  TANIGUCHI, Shunichiro
;  APPLICANT:  AMANO, Jun
;  APPLICANT:  YAZAWA, Kazuyuki
;  APPLICANT:  KANO, Yasunobu
;  APPLICANT:  NAKAMURA, Toshiyuki
;  APPLICANT:  SASAKI, Takayuki
;  TITLE OF INVENTION: Anaerobic bacterium as a drug for cancer gene therapy
;  FILE REFERENCE: 2001-WMC/01736
;  CURRENT APPLICATION NUMBER: US/09/816,391A
;  CURRENT FILING DATE:  2001-03-26
;  PRIOR APPLICATION NUMBER: JP 00/287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-09-816-391A-1

  Query Match             100.0%;  Score 192;  DB 9;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-54;
```

```
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 2
US-10-782-899-1
; Sequence 1, Application US/10782899
; Publication No. US20050025745A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, MINORU
;  APPLICANT:  TANIGUCHI, SHUNICHIRO
;  APPLICANT:  AMANO, JUN
;  APPLICANT:  YAZAWA, KAZUYUKI
;  APPLICANT:  KANO, YASUNOBU
;  APPLICANT:  NAKAMURA, TOSHIYUKI
;  APPLICANT:  SASAKI, TAKAYUKI
;  TITLE OF INVENTION: ANAEROBIC BACTERIUM AS A DRUG FOR CANCER GENE THERAPY
;  FILE REFERENCE: 671308-2001.1
;  CURRENT APPLICATION NUMBER: US/10/782,899
;  CURRENT FILING DATE:  2004-02-23
;  PRIOR APPLICATION NUMBER: 09/816,391
;  PRIOR FILING DATE: 2001-03-26
;  PRIOR APPLICATION NUMBER: JP 2000-287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-10-782-899-1

  Query Match             100.0%;  Score 192;  DB 21;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-54;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 3
US-10-470-565-1/c
```

```
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
;  APPLICANT: Societe des Produits Nestle S.A.
;  TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
;  FILE REFERENCE: 80290/WO
;  CURRENT APPLICATION NUMBER: US/10/470,565
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: EP 01102050.0
;  PRIOR FILING DATE: 2001-01-30
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 2256646
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
US-10-470-565-1

  Query Match             91.2%;  Score 175.2;  DB 19;  Length 2256646;
  Best Local Similarity   97.4%;  Pred. No. 9.2e-47;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2242172 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 2242113

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2242112 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 2242053

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2242052 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 2241993

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db    2241992 AAGAAGGATGCTTT 2241979


RESULT 4
US-10-425-115-60228/c
; Sequence 60228, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 60228
;   LENGTH: 425
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_15492C.1
US-10-425-115-60228

  Query Match             18.0%;  Score 34.6;  DB 20;  Length 425;
  Best Local Similarity   51.6%;  Pred. No. 0.47;
  Matches   79;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         16 CATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGG 75
              | |||| |    ||||||  ||||   |   |  |||  | |     |||||  |   |
Db        187 CTTGAAATCTACTGACAAAGATAAGACTAAATTTTTCACCCAAAACAAATACTATAATGC 128

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              ||  | |      ||  | | | ||| ||   ||| |  |  |  | || ||  ||   |
Db        127 AATGACAGCAATTAATACATTTGAAATGCCTATTTACATAATACAAAGCACTGCTAGACT 68
```

```
Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
              |  || |||||    |||| || ||| | | ||
Db         67 TACTGCCAACAATAGCTAAACAGAAGAGGTGGT 35


RESULT 5
US-10-027-632-167810/c
; Sequence 167810, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167810
;   LENGTH: 824
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-167810

  Query Match             17.3%;  Score 33.2;  DB 13;  Length 824;
  Best Local Similarity   59.6%;  Pred. No. 1.9;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              | ||| ||| |||||||   | |  | ||    |||   ||||   ||  |  | |||
Db        384 ATATAAATGAGAAAACCAAAAGAGTAAGCAATATTTAAAAGAAGGTTGAACATGGATCTA 325

Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGTC 169
              |||||| |||| |   |||||||  ||  |  ||
Db        324 TGATGAGAACAAGTCATAAGCAAGGGTTGTAATC 291


RESULT 6
US-10-027-632-167810/c
; Sequence 167810, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
```

```
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167810
;   LENGTH: 824
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-167810

  Query Match             17.3%;  Score 33.2;  DB 17;  Length 824;
  Best Local Similarity   59.6%;  Pred. No. 1.9;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              | ||| ||| |||||||   | |  | ||    |||   ||||   ||  |  | |||
Db        384 ATATAAATGAGAAAACCAAAAGAGTAAGCAATATTTAAAAGAAGGTTGAACATGGATCTA 325

Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGTC 169
              |||||| |||| |   |||||||  ||  |  ||
Db        324 TGATGAGAACAAGTCATAAGCAAGGGTTGTAATC 291


RESULT 7
US-10-027-632-6585
; Sequence 6585, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6585
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-6585

  Query Match             16.9%;  Score 32.4;  DB 13;  Length 484;
  Best Local Similarity   60.7%;  Pred. No. 2.8;
  Matches   51;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         36 ATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTT 95
              | ||| || | |    |  ||| | | ||| ||| |||| :|||||   ||| ||   |
Db        130 AGAATATTCTTTTTCACACCTACTGTTAATCCCTCCGGGTRAATAGGCATGACAATAATA 189

Qy         96 ATAAAACGCGGGTTTTCGCAGAAA 119
              || |||  |   |||||  |||||
Db        190 ATGAAATACTTTTTTTCTGAGAAA 213


RESULT 8
US-10-027-632-6585
```

```
; Sequence 6585, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6585
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-6585

  Query Match             16.9%;  Score 32.4;  DB 17;  Length 484;
  Best Local Similarity   60.7%;  Pred. No. 2.8;
  Matches   51;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         36 ATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTT 95
              | ||| || | |    |  ||| | | ||| ||| |||| :|||||   ||| ||   |
Db        130 AGAATATTCTTTTTCACACCTACTGTTAATCCCTCCGGGTRAATAGGCATGACAATAATA 189

Qy         96 ATAAAACGCGGGTTTTCGCAGAAA 119
              || |||  |   |||||  |||||
Db        190 ATGAAATACTTTTTTTCTGAGAAA 213


RESULT 9
US-10-398-221-8/c
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 495269
;   TYPE: DNA
;   ORGANISM: Listeria innocua
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(end)
;   OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8

  Query Match             16.7%;  Score 32;  DB 17;  Length 495269;
  Best Local Similarity   50.7%;  Pred. No. 1.1e+02;
```

```
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db     229177 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 229118

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               || |  |   |  | | || |||||     ||| | |   |||    |  || | | |
Db     229117 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 229058

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
              ||    ||    |   ||  | | ||  | ||
Db     229057 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 229026


RESULT 10
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
;   LENGTH: 3011208
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             16.7%;  Score 32;  DB 17;  Length 3011208;
  Best Local Similarity   50.7%;  Pred. No. 2.4e+02;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db    1529778 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 1529837

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               || |  |   |  | | || |||||     ||| | |   |||    |  || | | |
Db    1529838 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 1529897

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
              ||    ||    |   ||  | | ||  | ||
Db    1529898 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 1529929


RESULT 11
US-10-425-115-71957/c
; Sequence 71957, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 71957
;   LENGTH: 732
;   TYPE: DNA
```

```
;    ORGANISM: Zea mays
;    FEATURE:
;    NAME/KEY: unsure
;    LOCATION: (1)..(732)
;    OTHER INFORMATION: unsure at all n locations
;    FEATURE:
;    OTHER INFORMATION: Clone ID: MRT4577_165627C.1
US-10-425-115-71957

  Query Match             16.5%;  Score 31.6;  DB 20;  Length 732;
  Best Local Similarity   58.5%;  Pred. No. 6.3;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          54 TCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCG 113
               ||| || | |||    |     | ||||| ||||| |  ||||| ||| |  | | |  | |
Db         361 TCTTTTGTTAATTTGGTGCAACATATAGAGGTGAACAAGCTTATTAAAGGAAGATCCTTG 302

Qy         114 CAGAAACATGCGCTAGTATCATTGATGACAACAT 147
                | || || | | |  ||  || ||||| |  ||
Db         301 TACAATCAAGAGTTCATACAATAGATGAAACGAT 268


RESULT 12
US-10-674-124A-12097/c
; Sequence 12097, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 12097
;    LENGTH: 360
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    OTHER INFORMATION: AC007735.2_88547
;    FEATURE:
;    OTHER INFORMATION: Located on chromosome 7
;    FEATURE:
;    OTHER INFORMATION: Distance between a terminus base of telomere on
;    OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;    OTHER INFORMATION: sequence : 54730897
;    FEATURE:
;    OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;    OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;    OTHER INFORMATION: 5'-terminus of this base sequence : 7491
US-10-674-124A-12097

  Query Match             16.2%;  Score 31.2;  DB 19;  Length 360;
  Best Local Similarity   54.3%;  Pred. No. 6.1;
  Matches   63;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          47 TGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGG 106
               || |||   | |||||| || |||  | |||||||   | ||  |  |  | |||  ||
Db         313 TGCTTCCATTTTTTTCATTAACTTAAGAGAAATAGTATTCAATCCTTTGGTGAAATTCGA 254

Qy         107 GTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGT 162
                 | |  || |    ||      ||  | |||| |||| ||     ||  | ||||
```

```
Db         253 TGTCTACCAAATGTGTGTCTGTCTAGTACTGATTACAAGATTATAAAAATATAAGT 198

RESULT 13
US-09-764-891-8869/c
; Sequence 8869, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8869
;    LENGTH: 18272
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-764-891-8869

  Query Match             16.2%;  Score 31.2;  DB 10;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 41;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
               || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db       17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy         131 ATCATTGATGACAACATGGACTAA 154
               | ||   |||| |||| ||||| |
Db       17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 14
US-10-021-323-5754
; Sequence 5754, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
;  APPLICANT: Deikman, Jill
;  APPLICANT:  Feng, Paul C.C.
;  APPLICANT:  Fincher, Karen L.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(52274)B
;  CURRENT APPLICATION NUMBER: US/10/021,323
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: US 60/255, 619
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 5754
;    LENGTH: 466
;    TYPE: DNA
;    ORGANISM: Gossypium hirsutum
;    FEATURE:
;    OTHER INFORMATION: Clone ID: LIB3828-002-Q1-N6-F6
US-10-021-323-5754

  Query Match             15.9%;  Score 30.6;  DB 19;  Length 466;
  Best Local Similarity   58.1%;  Pred. No. 11;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
               | ||| ||   |   |  | || ||   | |||| | |   | | |  |  || ||||||
Db          27 ATATATATCATATGTCTTTCATTAACTTCTGGTTCTGGACTACATAATCCTTATTATCAT 86

Qy         136 TGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
               |||| | | |  |||  || ||| ||  |||||
Db          87 TGATCATACCCGGGAGGAACCAAGAGCCCTTGT 119
```

```
RESULT 15
US-10-260-238-5041/c
; Sequence 5041, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, Paul R.
;  APPLICANT:  Moughamer, Todd G.
;  APPLICANT:  Briggs, Steven P.
;  APPLICANT:  Cooper, Bret
;  APPLICANT:  Glazebrook, Jane
;  APPLICANT:  Goff, Stephen A.
;  APPLICANT:  Katagiri, Fumiyaki
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Provart, Nicholas
;  APPLICANT:  Ricke, Darrell
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;  FILE REFERENCE: 60111-NP
;  CURRENT APPLICATION NUMBER: US/10/260,238
;  CURRENT FILING DATE:  2002-09-26
;  PRIOR APPLICATION NUMBER: US 60/325,448
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/325,277
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/370,620
;  PRIOR FILING DATE: 2002-04-04
;  NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 5041
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Musa acuminata
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (12)..(13)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (23)..(23)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (416)..(416)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (547)..(547)
;   OTHER INFORMATION: n = any nucleotide
US-10-260-238-5041

  Query Match             15.9%;  Score 30.6;  DB 17;  Length 549;
  Best Local Similarity   51.9%;  Pred. No. 12;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         48 GATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGG 107
              || || | |   ||||| ||    |  |||| || | |  ||   |||    ||   | |
Db        235 GAATCATTTGAGTTCAAGACTGCTGAAGAAAAAGTTATCGAAGGTCTTGACCAAGCAGTG 176

Qy        108 TTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTG 167
               |  ||  |||  | |      |  | ||| ||||||      || |  |    | ||||
Db        175 ATGACGATGAATAAGGGAGAGATTGCTTTGGTGACAATTCCACCTGAATATGCCTTCTTG 116

Qy        168 TCCCCTGACCCAA 180
              ||| ||||  |||
Db        115 TCCACTGAATCAA 103


Search completed: August  6, 2005, 01:45:25
Job time : 428.986 secs
```

```
                          GenCore version 5.1.6
                 Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 23:28:19 ; Search time 102.28 Seconds
                                          (without alignments)
                                          3071.607 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:      2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                    %
Result            Query
  No.   Score     Match Length DB  ID                     Description
-----------------------------------------------------------------------
     1    32.6    17.0     278  4  US-09-513-999C-31601    Sequence 31601, A
c    2    31.2    16.2   22404  4  US-09-949-016-11765     Sequence 11765, A
c    3    31.2    16.2   22404  4  US-09-949-016-15398     Sequence 15398, A
c    4    31.2    16.2   46745  4  US-09-949-016-13964     Sequence 13964, A
     5    30.6    15.9     601  4  US-09-949-016-48509     Sequence 48509, A
c    6    30.6    15.9   58768  4  US-09-949-016-13175     Sequence 13175, A
     7    30.4    15.8    1569  4  US-09-710-279-603       Sequence 603, App
     8    30.4    15.8    3204  4  US-09-710-279-3485      Sequence 3485, Ap
c    9    29.8    15.5   10303  4  US-09-634-238-410       Sequence 410, App
c   10    29.8    15.5   99797  4  US-09-949-016-15255     Sequence 15255, A
c   11    29.6    15.4     601  4  US-09-949-016-187277    Sequence 187277,
c   12    29.6    15.4   36855  4  US-09-949-016-17095     Sequence 17095, A
    13    29.4    15.3     768  4  US-09-134-000C-3331     Sequence 3331, Ap
c   14    29.2    15.2     601  4  US-09-949-016-142351    Sequence 142351,
    15    29.2    15.2    2385  4  US-09-949-016-4010      Sequence 4010, Ap
    16    29.2    15.2    3077  4  US-09-461-912A-27       Sequence 27, Appl
    17    29.2    15.2    3092  1  US-08-426-627-3         Sequence 3, Appli
    18    29.2    15.2    3126  2  US-08-477-396A-3        Sequence 3, Appli
    19    29.2    15.2    3253  1  US-08-426-627-5         Sequence 5, Appli
    20    29.2    15.2    4623  4  US-09-949-016-15769     Sequence 15769, A
    21      29    15.1     601  4  US-09-949-016-44696     Sequence 44696, A
    22      29    15.1     832  4  US-09-621-976-2813      Sequence 2813, Ap
    23      29    15.1   36791  4  US-09-949-016-16390     Sequence 16390, A
    24      29    15.1   36791  4  US-09-949-016-16391     Sequence 16391, A
    25      29    15.1   42479  4  US-09-949-016-16631     Sequence 16631, A
    26      29    15.1   42479  4  US-09-949-016-16632     Sequence 16632, A
c   27      29    15.1   48940  4  US-09-949-016-16402     Sequence 16402, A
c   28      29    15.1   91062  4  US-09-949-016-13019     Sequence 13019, A
    29      29    15.1   92407  4  US-09-596-002-36        Sequence 36, Appl
c   30      29    15.1 1830121  4   US-09-557-884-1        Sequence 1, Appli
c   31      29    15.1 1830121  4   US-09-643-990A-1       Sequence 1, Appli
    32    28.8    15.0  636591  4  US-09-949-016-11808     Sequence 11808, A
    33    28.8    15.0  636591  4  US-09-949-016-13388     Sequence 13388, A
    34    28.6    14.9  189560  4  US-09-949-016-17202     Sequence 17202, A
c   35    28.6    14.9  263693  4  US-09-949-016-12386     Sequence 12386, A
c   36    28.6    14.9  263694  4  US-09-949-016-16915     Sequence 16915, A
    37    28.4    14.8   50381  4  US-09-949-016-17122     Sequence 17122, A
c   38    28.4    14.8   78125  4  US-09-949-016-16006     Sequence 16006, A
c   39    28.2    14.7   18264  4  US-09-949-016-16469     Sequence 16469, A
    40      28    14.6   61178  4  US-09-949-016-17369     Sequence 17369, A
    41      28    14.6  194537  4  US-09-949-016-12928     Sequence 12928, A
    42      28    14.6  201529  4  US-09-949-016-12740     Sequence 12740, A
c   43    27.8    14.5     730  3  US-08-943-731-128       Sequence 128, App
    44    27.8    14.5    1179  4  US-09-710-279-193       Sequence 193, App
    45    27.8    14.5    1257  1  US-08-330-154-1         Sequence 1, Appli

                                  ALIGNMENTS

RESULT 1
US-09-513-999C-31601
; Sequence 31601, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 31601
;   LENGTH: 278
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 152
;   OTHER INFORMATION: k=g or t
US-09-513-999C-31601

  Query Match             17.0%;  Score 32.6;  DB 4;  Length 278;
  Best Local Similarity   51.0%;  Pred. No. 0.11;
  Matches   77;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          18 TGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAA 77
               | | |||| |||       |  |   || | |     || | |||||||  || |||||
Db          99 TCATGTGGATTGTTTTCTGTCTTTAGGTATTAAAATGCTGTATTCAATAATTTKGGGGAC 158

Qy          78 ATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTG 137
               ||| |||| |   | | |||| |  |  |  | ||   |  |||| |   |    || ||
Db         159 ATATATGTCATTTCACATATATACAGATGTGTATCAGGGTGACATTCCAAAAGTGCACTG 218

Qy         138 ATGACAACATGGACTAAGCAAAAGTGCTTGT 168
                |||      ||  || | |   ||  || |
Db         219 CTGAGTTAGAGGTATATGTATTTGTAATTTT 249


RESULT 2
US-09-949-016-11765/c
; Sequence 11765, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
```

```
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11765
;   LENGTH: 22404
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(22404)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11765

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 22404;
  Best Local Similarity   60.7%;  Pred. No. 3.1;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      19171 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 19112

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      19111 AACACCCATGAGAACAGGGACTCA 19088


RESULT 3
US-09-949-016-15398/c
; Sequence 15398, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15398
;   LENGTH: 22404
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(22404)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15398

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 22404;
  Best Local Similarity   60.7%;  Pred. No. 3.1;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      19171 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 19112
```

```
Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      19111 AACACCCATGAGAACAGGGACTCA 19088


RESULT 4
US-09-949-016-13964/c
; Sequence 13964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13964
;   LENGTH: 46745
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(46745)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13964

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 46745;
  Best Local Similarity   54.3%;  Pred. No. 4.5;
  Matches   63;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy         46 CTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCG 105
              ||| |||  ||  | |||  ||||| || |  ||| | ||    ||| || ||      |
Db      40974 CTGCTTCCACTGCTGTCATGACCTTAGGTGGCATAAAAGTCCTTACCTTTCTACCCTAAG 40915

Qy        106 GGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
              | |   |  |    | |    ||||||| || |  ||| |   |  ||||  ||||
Db      40914 GATGGGCCAATTCTCTTCTTTTAGTATCTTTCAGAACAGCTGTGCTTAAGTTAAAG 40859


RESULT 5
US-09-949-016-48509
; Sequence 48509, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48509
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-48509

  Query Match             15.9%;  Score 30.6;  DB 4;  Length 601;
```

```
  Best Local Similarity   58.1%;  Pred. No. 0.86;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         56 TATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCA 115
              || ||||| || |||  || |  |  | | |||||     ||||| |     | ||  |
Db        149 TACTTTCATTAACTTTAGGAAGCTGAAGGGGAAAAAATCAATAAAGCAGCTCTCTTGACT 208

Qy        116 GAAACATGCGCTAGTATCATTGATGACAACATG 148
               ||| ||| |||   || ||  ||||  | |||
Db        209 CAAAGATGAGCTGAAATAATACATGAATAAATG 241


RESULT 6
US-09-949-016-13175/c
; Sequence 13175, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13175
;   LENGTH: 58768
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(58768)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13175

  Query Match             15.9%;  Score 30.6;  DB 4;  Length 58768;
  Best Local Similarity   58.1%;  Pred. No. 8.2;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         56 TATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCA 115
              || ||||| || |||  || |  |  | | |||||     ||||| |     | ||  |
Db      43027 TACTTTCATTAACTTTAGGAAGCTGAAGGGGAAAAAATCAATAAAGCAGCTCTCTTGACT 42968

Qy        116 GAAACATGCGCTAGTATCATTGATGACAACATG 148
               ||| ||| |||   || ||  ||||  | |||
Db      42967 CAAAGATGAGCTGAAATAATACATGAATAAATG 42935


RESULT 7
US-09-710-279-603
; Sequence 603, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 603
;   LENGTH: 1569
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-603

  Query Match             15.8%;  Score 30.4;  DB 4;  Length 1569;
  Best Local Similarity   57.3%;  Pred. No. 1.6;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              ||  |  |||   ||| |   |  | ||   | || |   |   ||||||||||    ||
Db         87 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 146

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              |  ||| || | ||  | |||| | ||| |||| ||
Db        147 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 182


RESULT 8
US-09-710-279-3485
; Sequence 3485, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3485
;   LENGTH: 3204
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3485

  Query Match             15.8%;  Score 30.4;  DB 4;  Length 3204;
  Best Local Similarity   57.3%;  Pred. No. 2.3;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              ||  |  |||   ||| |   |  | ||   | || |   |   ||||||||||    ||
Db        542 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 601

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              |  ||| || | ||  | |||| | ||| |||| ||
Db        602 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 637


RESULT 9
US-09-634-238-410/c
; Sequence 410, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
;  APPLICANT:  Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka J.
;  APPLICANT:  Bloksberg, Leonard, N.
;  APPLICANT:  Lubbers, Mark W.
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna C.
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul W.
;  APPLICANT:  Reid, Julian R.
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION:  Polynucleotides, materials incorporating
;  TITLE OF INVENTION:  them and methods for using them.
;  FILE REFERENCE:  11000.1043U1
;  CURRENT APPLICATION NUMBER: US/09/634,238
;  CURRENT FILING DATE:  2000-08-08
```

```
;  NUMBER OF SEQ ID NOS:  422
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 410
;   LENGTH: 10303
;   TYPE: DNA
;   ORGANISM: Lactobacillus rhamnosus
US-09-634-238-410

  Query Match             15.5%;  Score 29.8;  DB 4;  Length 10303;
  Best Local Similarity   60.5%;  Pred. No. 6.7;
  Matches   49;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        112 CGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCC 171
              || |||||   ||      ||  |||| ||   |||| |  | |||||||   ||||| |
Db       8088 CGAAGAAAGCAGCAAGGACATTGTTGAAGAAGTCATGCAGCAGGCAAAAGAAATTGTCGC 8029

Qy        172 CTGACCCAAGAAGGATGCTTT 192
              | |  | || | ||  |||||
Db       8028 CGGCTCGAAAATGGCGGCTTT 8008


RESULT 10
US-09-949-016-15255/c
; Sequence 15255, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15255
;   LENGTH: 99797
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99797)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15255

  Query Match             15.5%;  Score 29.8;  DB 4;  Length 99797;
  Best Local Similarity   51.1%;  Pred. No. 21;
  Matches   70;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         40 TCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAA 99
              | | |  ||| || |  ||||       ||   |  |||  |||||| |     ||| |
Db      32501 TGTAGGATGAATCCTGCATTTGATGAGACTAGTGTTAAAAGGATGTGCAGGGGGTTAGAC 32442

Qy        100 AACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAA 159
                 |||||     |||| |||| |  |   || ||      | ||||        | ||
Db      32441 GGGGCGGGGGAGAGCAGTAACAGGGACAGTTACCAAGCCATAAAACAAACGAGCTGAAAC 32382

Qy        160 AGTGCTTGTCCCCTGAC 176
              | | |||| || |||||
Db      32381 AATTCTTGCCCTCTGAC 32365


RESULT 11
US-09-949-016-187277/c
; Sequence 187277, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
```

```
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187277
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-187277

  Query Match             15.4%;  Score 29.6;  DB 4;  Length 601;
  Best Local Similarity   59.5%;  Pred. No. 2;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         96 ATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAG 155
              |||  || |  ||  ||   ||||  | |     |||| || ||||||| |||  | |||
Db        108 ATATGACCCCAGTAGTCTTTGAAAGCTCCTTCCCTATCTTTTATGACAAGATGTTCCAAG 49

Qy        156 CAAAAGTGCTTGTCCCCTGACCCA 179
              |  |  ||  | |  |||| ||||
Db         48 CTTATTTGTATATTTCCTGCCCCA 25


RESULT 12
US-09-949-016-17095/c
; Sequence 17095, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17095
;   LENGTH: 36855
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17095

  Query Match             15.4%;  Score 29.6;  DB 4;  Length 36855;
  Best Local Similarity   59.5%;  Pred. No. 15;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         96 ATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAG 155
              |||  || |  ||  ||   ||||  | |     |||| || ||||||| |||  | |||
Db      29826 ATATGACCCCAGTAGTCTTTGAAAGCTCCTTCCCTATCTTTTATGACAAGATGTTCCAAG 29767

Qy        156 CAAAAGTGCTTGTCCCCTGACCCA 179
              |  |  ||  | |  |||| ||||
Db      29766 CTTATTTGTATATTTCCTGCCCCA 29743


RESULT 13
US-09-134-000C-3331
; Sequence 3331, Application US/09134000C
```

```
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3331
;   LENGTH: 768
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-3331

  Query Match             15.3%;  Score 29.4;  DB 4;  Length 768;
  Best Local Similarity   63.4%;  Pred. No. 2.6;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         97 TAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGC 156
              | |  ||  ||||  |||| ||||| | |  | || ||||||| ||   |   |  ||
Db        327 TTATGCGACGGTTGGCGCAAAAACAAGTGTCAATAACATTGATTACTTTAAAAATGAACT 386

Qy        157 AAAAGTGCTTG 167
              ||||| |||||
Db        387 AAAAGAGCTTG 397


RESULT 14
US-09-949-016-142351/c
; Sequence 142351, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 142351
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-142351

  Query Match             15.2%;  Score 29.2;  DB 4;  Length 601;
  Best Local Similarity   54.7%;  Pred. No. 2.7;
  Matches   58;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         55 CTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGC 114
              ||| | | || |   ||    ||| |||  |||  |    ||||||| | ||||| || |
Db        265 CTAGTGTTAAAAAACTCACACAAAAAGAGCTGAGGAAGGCTATAAAAAGAGGGTTCTCAC 206

Qy        115 AGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAA 160
              |     ||||   |  |  |    |  ||  |  ||||  ||||||
Db        205 ACTTGGATGCCTAATAACAAAAACTATCACAAAAGACTCTGCAAAA 160


RESULT 15
US-09-949-016-4010
; Sequence 4010, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4010
;   LENGTH: 2385
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-4010

  Query Match             15.2%;  Score 29.2;  DB 4;  Length 2385;
  Best Local Similarity   50.0%;  Pred. No. 5.4;
  Matches   73;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy         40 TCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAA 99
              || ||     || | ||| | |   ||  | |      |||| | ||  ||   | | ||
Db       1281 TCATGAAGATTTGGCCTACTCTTCTTATGTCCACCAGTATAGCTCTGGGAAATTTGACAA 1340

Qy        100 AACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAA 159
                | |   ||   ||||||| |     | ||||||||    ||| | |     |||   |
Db       1341 TGCACTATTTCATGCAGAAAGAGCTATTGGTATCATTACCCACATCCTACCTGAAGATCA 1400

Qy        160 AGTGCTTGTCCCCTGACCCAAGAAGG 185
                | ||| |  | |   | |||| ||
Db       1401 TCTTCTTTTGGCTTCTTCAAAGAGGG 1426


Search completed: August  6, 2005, 01:33:31
Job time : 107.28 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         August  5, 2005, 22:54:59 ; Search time 1949.31 Seconds
                                           (without alignments)
                                           3749.196 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:      68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                      %
Result                Query
   No.     Score    Match  Length  DB  ID                      Description
-----------------------------------------------------------------------------
     1      40.8     21.2     673   9  CE144436                CE144436 tigr-gss-
c    2      36.4     19.0     749   8  BH039236                BH039236 RPCI-24-2
     3      34.6     18.0     852   4  BG542199                BG542199 602571637
     4      34.4     17.9     590   8  BH378221                BH378221 AG-ND-119
c    5      34.2     17.8     600   8  AZ964976                AZ964976 2M0234D24
     6      34.2     17.8     619   5  BP519798                BP519798 BP519798
     7      34.2     17.8     878   7  CR281388                CR281388 CR281388
c    8      33.8     17.6     945   9  CNS079HE                AL435288 T3 end of
c    9      33.6     17.5     845   9  CNS0757U                AL429760 clone BA0
    10      33.4     17.4     494   6  CA394392                CA394392 cs51a07.y
    11      33.4     17.4     728   8  BH885133                BH885133 hw61c11.b
    12      33.2     17.3     449   7  CN673288                CN673288 A0932B01-
    13      33.2     17.3     467   7  CF166270                CF166270 B0768E01-
    14      33.2     17.3     493   7  CK391130                CK391130 K0827F01-
    15      33.2     17.3     501   6  CD552026                CD552026 B0337C08-
    16      33.2     17.3     501   7  CF165236                CF165236 B0753H06-
    17      33.2     17.3     503   7  CF168303                CF168303 B0797D04-
    18      33.2     17.3     538   7  CF162296                CF162296 B0709D06-
    19      33.2     17.3     551   7  CK391037                CK391037 K0826D11-
    20      33.2     17.3     552   7  CK390292                CK390292 K0817C11-
    21      33.2     17.3     552   7  CN690065                CN690065 E0288A10-
    22      33.2     17.3     555   6  CA885628                CA885628 B0120F01-
    23      33.2     17.3     555   6  CD553021                CD553021 B0353C01-
    24      33.2     17.3     557   6  CD553973                CD553973 B0369F09-
    25      33.2     17.3     572   7  CN682591                CN682591 E0166H03-
    26      33.2     17.3     579   6  CD540845                CD540845 B0226C02-
    27      33.2     17.3     589   7  CN676248                CN676248 A0977D08-
    28      33.2     17.3     596   7  CN681180                CN681180 E0147C03-
c   29      33.2     17.3     610   8  CC007613                CC007613 PUEDB53TD
c   30      33.2     17.3     611   8  AQ017410                AQ017410 CIT-HSP-2
    31      33.2     17.3     613   6  CD543738                CD543738 B0251E12-
    32      33.2     17.3     616   7  CF898187                CF898187 A0235E06-
    33      33.2     17.3     621   6  CD541777                CD541777 B0235A12-
    34      33.2     17.3     623   6  CA883270                CA883270 B0102B10-
    35      33.2     17.3     630   8  AQ268787                AQ268787 RPCI11-74
    36      33.2     17.3     642   2  BB658280                BB658280 BB658280
    37      33.2     17.3     663   7  CF904613                CF904613 A0424D01-
    38      33.2     17.3     771   7  CN455104                CN455104 UI-M-HN0-
    39      33.2     17.3     774   7  CO433921                CO433921 UI-M-HX0-
    40      33.2     17.3     777   7  CN459909                CN459909 UI-M-HB0-
    41      33.2     17.3     783   6  CB524602                CB524602 UI-M-FY0-
    42      33.2     17.3     806   7  CF739778                CF739778 UI-M-HD0-
    43      33.2     17.3    3083   3  AK035944                AK035944 Mus muscu
    44      33.2     17.3    3969   9  AY414503                AY414503 Mus muscu
    45      33.2     17.3    4526   3  AK031367                AK031367 Mus muscu


                                  ALIGNMENTS


RESULT 1
CE144436
LOCUS       CE144436                 673 bp    DNA     linear   GSS 25-SEP-2003
DEFINITION  tigr-gss-dog-17000371273067 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE144436
VERSION     CE144436.1  GI:35257693
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..673
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             21.2%;  Score 40.8;  DB 9;  Length 673;
  Best Local Similarity   56.8%;  Pred. No. 0.082;
  Matches   75;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         23 TGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGA 82
              ||||||| |   ||| || |  || | | | ||| |||||||||   |  |  |||||
Db        326 TGGCTTGTCCTTCATTATGTCCTCAGCTCCTTCTGTTTTCAATAGAGTTTGTCAAATAAG 385
```

```
Qy         83 TGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGAC 142
              |   || |   | ||||   |  ||||||  ||  |  |  ||  | |  ||| |||
Db        386 TAGCAAGAGTTTGATAAGCAGGAGGTTTTAGCTTAGTCTGGCTATCGGTTCAGTGACTGT 445

Qy        143 AACATGGACTAA 154
              |   || | |||
Db        446 ATTCTGTAGTAA 457


RESULT 2
BH039236/c
LOCUS       BH039236                 749 bp    DNA     linear   GSS 17-JUL-2001
DEFINITION  RPCI-24-254N24.TV RPCI-24 Mus musculus genomic clone
            RPCI-24-254N24, genomic survey sequence.
ACCESSION   BH039236
VERSION     BH039236.1  GI:14816964
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-254N24.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 254  row: N  column: 24
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..749
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-254N24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match             19.0%;  Score 36.4;  DB 8;  Length 749;
  Best Local Similarity   54.5%;  Pred. No. 2.1;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         13 GGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCG 72
              |||  |   |||||||| ||| ||| |    || |||||     | ||  || || |||
Db        456 GGCATTCTGGTGGCTTCACATGCAAATCACTGCCTGATGATAATGTTGCCAGTATCTTGC 397

Qy         73 GGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTAT 132
                | | | | |||||| ||| | ||  |       |       ||| |   ||||  |||
Db        396 ATGCATTGGGTGTGAATACCATAATGCAGTAGCCCTCCATTTTGAAGCCATCGCTGTTAT 337

Qy        133 CATTGATGACAACA 146
```

```
               | |||| ||| |
Db        336 CTCTGATTGCAAAA 323


RESULT 3
BG542199
LOCUS       BG542199                 852 bp    mRNA    linear   EST 03-APR-2001
DEFINITION  602571637F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4696041 5',
            mRNA sequence.
ACCESSION   BG542199
VERSION     BG542199.1  GI:13534432
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1522  row: j  column: 10
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1..852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4696041"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             18.0%;  Score 34.6;  DB 4;  Length 852;
  Best Local Similarity   55.4%;  Pred. No. 7.9;
  Matches   67;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         38 AATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTAT 97
              | | | |||| ||| | | |      |   | |||||||  ||  ||   |  ||  |||
Db        446 ACTATGGTCTCATTGGCCAACGCCACAATACATCGGGGACCTAAGTGACCAGCCCAATAT 505

Qy         98 AAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCA 157
               ||  ||||| | |  |  | ||  |  | |||  |||| |    ||||||||||  |
Db        506 CAACGGCGGGCTCTATCCTATACTCGGACAAGTCACATTTTTCCATACATGGACTACCCC 565

Qy        158 A 158
              |
Db        566 A 566


RESULT 4
BH378221
LOCUS       BH378221                 590 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-119J16.TR ND-TAM Anopheles gambiae genomic clone
            AG-ND-119J16, genomic survey sequence.
```

```
ACCESSION   BH378221
VERSION     BH378221.1  GI:17324363
KEYWORDS    GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
            Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
            Gardner,M.J. and Collins,F.H.
  TITLE     Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
  JOURNAL   Mol. Genet. Genomics 268 (6), 720-728 (2003)
  MEDLINE   22542063
   PUBMED   12655398
COMMENT     Other_GSSs: AG-ND-119J16.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Anopheles gambiae"
                     /mol_type="genomic DNA"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-119J16"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
ORIGIN

  Query Match             17.9%;  Score 34.4;  DB 8;  Length 590;
  Best Local Similarity   53.8%;  Pred. No. 8.4;
  Matches   71;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          38 AATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTAT 97
               ||| || ||| ||   ||||| ||| |   || | |    ||     | ||      | |
Db         459 AATTTTATCTTATCTTTCTATCTTCCACTTCTACCGAAGCATTAGAATTAATTTGGATGT 518

Qy          98 AAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCA 157
               || |  |  ||||| |  ||| | |  ||| || |||   |||| ||    |||  |||
Db         519 AACAGACCAGTTTTAGTTGAATCTTAAGCTCGTCTCAACAATGAAAAACGTGACGCAGCT 578

Qy         158 AAAGTGCTTGTC 169
                ||   || | |
Db         579 CAACCCCTAGCC 590


RESULT 5
AZ964976/c
LOCUS       AZ964976                 600 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0234D24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0234D24 R, genomic survey sequence.
ACCESSION   AZ964976
VERSION     AZ964976.1  GI:13836203
KEYWORDS    GSS.
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0234  row: D  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0234D24"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 8;  Length 600;
  Best Local Similarity   56.8%;  Pred. No. 9.8;
  Matches   63;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          51 TCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTT 110
               | | |||| | ||| ||||  ||  | |  | |   | || ||||||   ||    ||
Db         330 TAGCCTATCTGCAAAACCTCAGGATACAGGGGTTGAAGAAGCCTTATTTCACAACTGTCC 271

Qy         111 TCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               |  || | | | ||    || ||||||   | |||| ||| ||   ||| |
Db         270 TAACATATATAAGCCTATGTCTCATTGTGCATAACAAGGAGTATATAAAGG 220


RESULT 6
BP519798
LOCUS       BP519798                 619 bp    mRNA    linear   EST 17-SEP-2003
DEFINITION  BP519798 Hydra magnipapillata cDNA library Hydra magnipapillata
```

```
            cDNA clone hmp_17065, mRNA sequence.
ACCESSION   BP519798
VERSION     BP519798.1  GI:34823167
KEYWORDS    EST.
SOURCE      Hydra magnipapillata
  ORGANISM  Hydra magnipapillata
            Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
            Hydridae; Hydra.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Hayakawa,S., Hwang,J.S., Nishimiya-Fujisawa,C., Ogura,A., Ikeo,K.,
            Fujisawa,T. and Gojobori,T.
  TITLE     Hydra EST project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jung Shan Hwang
            Center for Information Biology and DNA Data Bank of Japan
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-55-981-6847(ex.6898)
            Fax: 81-55-981-6848
            Email: jhwang@lab.nig.ac.jp, URL:http://www.cib.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..619
                     /organism="Hydra magnipapillata"
                     /mol_type="mRNA"
                     /strain="105"
                     /db_xref="taxon:6085"
                     /clone="hmp_17065"
                     /tissue_type="whole body"
                     /dev_stage="adult budding stage"
                     /clone_lib="Hydra magnipapillata cDNA library"
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 5;  Length 619;
  Best Local Similarity   50.9%;  Pred. No. 9.9;
  Matches   81;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         17 ATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGA 76
              |||||||||  |  |||  |  | | |     |     |   | |||| |   |     |
Db         30 ATGAAGTGGGCTTTCAAAGAAGAACATACTATAGAAAGCAGATGTCAAGAGTCTACAAAA 89

Qy         77 AATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATT 136
              |  ||| |  |  | ||  ||  ||  |  ||| | |  |||| |    || ||  ||
Db         90 ATAAGAAGCAAGTATCCGGATCGAATTCCTGTTGTAGTTGAAAAAGCTCCTCGTTCCACA 149

Qy        137 GATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGA 175
                | |  | || ||| ||  |||  | ||||| |||| |
Db        150 ATTCAAGATATTGACAAAAGAAAGTTTCTTGTTCCCTCA 188


RESULT 7
CR281388
LOCUS       CR281388                 878 bp    mRNA    linear   EST 27-FEB-2004
DEFINITION  CR281388 Oryza sativa library (Han B) Oryza sativa cDNA clone
            y620d12p5, mRNA sequence.
ACCESSION   CR281388
VERSION     CR281388.1  GI:44667954
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Ying,K., Li,Y., Guan,J.P., Zhu,J.J.,
            Zhao,Q., Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L.,
            Weng,Q.J., Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T.,
            Zhang,Y.J., Lu,Y., Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X.,
            Zhang,L., Lan,L.F., Chen,W., Wu,S.A. and Xue,Y.B.
  TITLE     Rice cDNA EST clone
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Han Bin
            National Center for Gene Research
            Chinese Academy of Sciences
            500# Cao Bao Road, Shanghai 200233, China
            Email: bhan@ncgr.ac.cn
            Clone requests: bhan@ncgr.ac.cn
             This is rice cdna est clone
            Web site: http://www.ncgr.ac.cn.
FEATURES             Location/Qualifiers
     source          1..878
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="y620d12p5"
                     /clone_lib="Oryza sativa library (Han B)"
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 7;  Length 878;
  Best Local Similarity   49.2%;  Pred. No. 11;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          2 CTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTT 61
              |||| |   |   |||||| |  |  |    |  | ||    |  | | |  | |||
Db        526 CTGGTCATTGTACCCATGAGGGTGACTATTGACGAGAAGTCGGGTTTACTGTTGTATCAG 585

Qy         62 CAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACA 121
               | |    ||  ||| | |  | | ||||        || |||||  ||||||  ||
Db        586 GATTTGGGTCTTGGATAAAATTCTTAAAATTTGGGACAAGCGCGGTATTTCGCTCAATGG 645

Qy        122 TGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAAG 181
                | || |||| |    ||| |   |  || | || | ||    |  ||||||||   |
Db        646 CCCTCTGGTATAAAATCTGAGACACTTTACAACGCTACAGAAGGTACCCCCTGACAGCAA 705

Qy        182 AAG 184
               ||
Db        706 TAG 708


RESULT 8
CNS079HE/c
LOCUS       CNS079HE                 945 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T3 end of clone BB0AA029A02 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL435288
VERSION     AL435288.1  GI:12218701
KEYWORDS    GSS.
SOURCE      Pichia angusta
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 945)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
   PUBMED   11152888
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
```

```
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1..945
                     /organism="Pichia angusta"
                     /mol_type="genomic DNA"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA029A02"
                     /clone_lib="BB0AA"
                     /note="end : T3"
     misc_feature    <4..>944
                     /note="similar to Saccharomyces cerevisiae ORF YMR290c [
                     HAS1 ; helicase associated with Set1p ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
ORIGIN

  Query Match             17.6%;  Score 33.8;  DB 9;  Length 945;
  Best Local Similarity   56.9%;  Pred. No. 15;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              || |||  |||||| |||||||||  |||||  ||||  |    || |  | ||  || |
Db        387 CAGAATTCTGTCTGCTTCGTCTATAATCAATGTCTTCAAGTTCTTAAAGATAAATCCCTT 328

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACA 143
                ||    ||  ||  ||    || |   |   |||  || |||| | |
Db        327 CGTATTTTGCAAGTGATCCAGCAATCTACCTGGAGTGGCAATGATCAAA 279


RESULT 9
CNS0757U/c
LOCUS       CNS0757U                 845 bp    DNA     linear   GSS 07-JUL-2001
DEFINITION  clone BA0AB033E09 of library BA0AB from strain CLIB 210 of
            Kluyveromyces lactis, genomic survey sequence.
ACCESSION   AL429760
VERSION     AL429760.1  GI:12212954
KEYWORDS    GSS.
SOURCE      Kluyveromyces lactis
  ORGANISM  Kluyveromyces lactis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 845)
  AUTHORS   Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
            Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 11.
            Kluyveromyces lactis
  JOURNAL   FEBS Lett. 487 (1), 66-70 (2000)
  MEDLINE   20584721
   PUBMED   11152886
REFERENCE   3  (bases 1 to 845)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1..845
                     /organism="Kluyveromyces lactis"
                     /mol_type="genomic DNA"
                     /strain="CLIB 210"
                     /variety="lactis"
                     /db_xref="taxon:28985"
                     /clone="BA0AB033E09"
                     /clone_lib="BA0AB"
     misc_feature    <1..>845
                     /note="similar to Saccharomyces cerevisiae ORF YPR010c [
                     RPA135 ; DNA-directed RNA polymerase I, 135 KD subunit ]"
                     /evidence=not_experimental
ORIGIN

  Query Match             17.5%;  Score 33.6;  DB 9;  Length 845;
  Best Local Similarity   54.1%;  Pred. No. 16;
  Matches   66;  Conservative    1;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         23 TGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGA 82
              | | |||| | |  | |||| |||| |    |||| | | |||||| |   || | |||
Db        165 TAGTTTGAGATGTTTGATCTGGTCTAACGGATCTAATCTGAATACCATATTGGGAGTAGG 106

Qy         83 TGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGAC 142
               |  |   |  |||  ||| |  ||| |    | |  |||  | | || ||   :   ||
Db        105 AGGCACCTCTGTTAGCAAATGAAGGTCTAATGATAGCCATTGGATGGTTTCTACKTGCAC 46

Qy        143 AA 144
              ||
Db         45 AA 44


RESULT 10
CA394392
LOCUS       CA394392                 494 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  cs51a07.y1 Human Retinal pigment epithelium/choroid cDNA
            (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs51a07
            5', mRNA sequence.
ACCESSION   CA394392
VERSION     CA394392.1  GI:24728898
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
            Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
  TITLE     Expressed sequence tag analysis of human RPE/choroid for the
            NEIBank Project: Over 6000 non-redundant transcripts, novel genes
            and splice variants
  JOURNAL   Mol. Vis. 8 (4), 205-220 (2002)
  MEDLINE   22103460
   PUBMED   12107410
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
```

```
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 51  row: a  column: 07
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1..494
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="cs51a07"
                     /tissue_type="RPE/choroid"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human Retinal pigment epithelium/choroid cDNA
                     (Un-normalized, unamplified): cs"
                     /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
                     eyes (75-80 years old) yielded approximately 600 mg of
                     dissected RPE/choroid tissue.  This in turn yielded 340 ug
                     of total RNA and 7 ug of mRNA. A directionally cloned cDNA
                     library in the pCMVSPORT6 vector was constructed at Life
                     Technologies (Rockville, MD; now part of Invitrogen Corp),
                     essentially following the protocols of the SuperScript
                     Plasmid System (Invitrogen Corp.
                     <http://www.invitrogen.com/>). The library code
                     designation was cs. For this library, cDNA inserts were
                     cloned into the NotI/MluI sites of the vector. EST
                     analysis was performed on the unamplified library at the
                     NIH Intramural Sequencing Center (NISC)."
ORIGIN

  Query Match             17.4%;  Score 33.4;  DB 6;  Length 494;
  Best Local Similarity   52.5%;  Pred. No. 17;
  Matches   73;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         42 TTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAA 101
              || |||  || | ||  | | |  | || |  | ||||  || ||   | |   || ||
Db          7 TTTTCTACTTTGCCTCCTCTTACAAGCTACAAGCAAATGAATCTGCTCAGCAGAATGAAG 66

Qy        102 CGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
              |     |||  ||| ||| ||    |   || |||  |    | ||   |||||||  |
Db         67 CCTATTTTTCAGCATAAATATTGCTTGCCATGATTAAAATAGAAATCAGCTAAGCATGAT 126

Qy        162 TGCTTGTCCCCTGACCCAA 180
              | |||  |   ||  ||||
Db        127 TTCTTTACTGGTGGACCAA 145


RESULT 11
BH885133
LOCUS       BH885133                 728 bp    DNA     linear   GSS 05-AUG-2002
DEFINITION  hw61c11.b1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone hw61c11 5', genomic survey sequence.
ACCESSION   BH885133
VERSION     BH885133.1  GI:22121030
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: hw61  row: c  column: 11
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 728.
FEATURES             Location/Qualifiers
     source          1..728
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="hw61c11"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by  fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end  repaired, adaptor ligated and size
                     fractionated using sephadex. The  resulting fragments were
                     between 0.8 and 3 kb and were cloned into  the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
ORIGIN

  Query Match             17.4%;  Score 33.4;  DB 8;  Length 728;
  Best Local Similarity   62.7%;  Pred. No. 18;
  Matches   52;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        110 TTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTC 169
              |||  ||| ||||||| || || ||  |||  |||| | | | |||   |  || ||  ||
Db        171 TTCATAGAGACATGCGATAATAGCAGGGATAACAATAAGAAATAATTCATCGTCCTCTTC 230

Qy        170 CCCTGACCCAAGAAGGATGCTTT 192
              | |     |||| ||||  || |
Db        231 CTCCTCTTCAAGCAGGAACCTAT 253


RESULT 12
CN673288
LOCUS       CN673288                 449 bp    mRNA    linear   EST 17-MAY-2004
DEFINITION  A0932B01-5 NIA Mouse Embryonic Stem (ES) cell (Lif+, 48 h, high
            density) cDNA library (Long) Mus musculus cDNA clone NIA:A0932B01
            IMAGE:30769836 5', mRNA sequence.
ACCESSION   CN673288
VERSION     CN673288.1  GI:47439739
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y.,
            VanBuren,V., Falco,G., Martin,P.R., Stagg,C.A., Bassey,U.C.,
            Wang,Y., Carter,M.G., Hamatani,T., Aiba,K., Akutsu,H., Sharova,L.,
            Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S.,
            Nagaraja,R., Boheler,K.R., Taub,D., Hodes,R.J., Longo,D.L.,
            Schlessinger,D., Keller,J., Klotz,E., Kelsoe,G., Umezawa,A.,
            Vescovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A.,
            D'Urso,M., Kelso,J., Hide,W. and Ko,M.S.
  TITLE     Transcriptome analysis of mouse stem cells and early embryos
  JOURNAL   PLoS Biol. 1 (3), 410-419 (2003)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: A0932  row: B  column: 01
            Seq primer: M13 Reverse
            High quality sequence stop: 449
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Mus musculus"
```

```
                     /mol_type="mRNA"
                     /strain="129Sv/EvTac"
                     /db_xref="niaEST:A0932B01-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:A0932B01 IMAGE:30769836"
                     /sex="Male"
                     /tissue_type="Embryonic Stem Cell"
                     /cell_line="129.3 ES cells"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Embryonic Stem (ES) cell (Lif+, 48
                     h, high density) cDNA library (Long)"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
                     Site_2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library (Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). ES
                     cells were plated at density 3x104/cm2, on gelatin-coated
                     plates and cultured for 48 hrs at 37 OC, 5% CO2. Culture
                     medium: DMEM supplemented with 15% FBS, 2 mM L-glutamine,
                     0.1 mM NEAA, 1mM Sodium pyruvate, 0.1 mM
                     beta-mercaptoethanol, 1000 U/ml LIF, 100 U/ml penicillin,
                     and 100 ug/ml streptomycin. Double-stranded cDNAs were
                     synthesized with an Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     2ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
                     vector. The DH10B E. coli host was transformed with the
                     ligation mixture by the standard chemical method. The
                     average insert size is about 2.7 kb. The library was
                     constructed by Yulan Piao."
ORIGIN

  Query Match             17.3%;  Score 33.2;  DB 7;  Length 449;
  Best Local Similarity   52.1%;  Pred. No. 19;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
              |||||   ||| ||||  ||  | |  |     ||  |  |   |  | |  | ||||||
Db        274 AAGTGTGGTGATAAGCCAAACATGGATTCCACAGTGAAACTAGGAGCCGTGGGTGGAAAT 333

Qy         80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGAT 139
               ||| | ||   |||  || ||| |    ||| | | | | || | | | |    | |
Db        334 GGATTTAAAGTTCCTCTTACAACACCTCATTTGGAAAAGAGATACTCGATTCAGTTCGGG 393

Qy        140 GACAACATGGACTAAGCAAAAG 161
               ||||    ||  ||| | |||
Db        394 AACAATGAAGAAGAAGAAGAAG 415


RESULT 13
CF166270
LOCUS       CF166270                 467 bp    mRNA    linear   EST 25-JUL-2003
DEFINITION  B0768E01-5 NIA Mouse Embryonic Germ Cell cDNA Library (Long) Mus
            musculus cDNA clone NIA:B0768E01 IMAGE:30464400 5', mRNA sequence.
ACCESSION   CF166270
VERSION     CF166270.1  GI:33275824
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
  TITLE     Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
  JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
  MEDLINE   21429098
   PUBMED   11544199
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0768  row: E  column: 01
            Seq primer: M13 Reverse
            High quality sequence stop: 467
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..467
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:B0768E01-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0768E01 IMAGE:30464400"
                     /sex="male"
                     /dev_stage="embryonic day 8"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Embryonic Germ Cell cDNA Library
                     (Long)"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
                     Site_2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library (Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
                     RNAs were obtained from Dr.Mark G. Carter (NIH/NIA-IRP).
                     EG cells were cultured at 37. C, 5% CO2 in DMEM
                     supplemented with 15% ES cell-qualified FBS, 0.1mM
                     non-essential amino acids, 2 mM glutamine,
                     penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM
                     beta-mercaptoethanol, and 10000000 units of LIF per liter.
                     Double-stranded cDNAs were synthesized with an Oligo(dT)
                     primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     2.5 ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
                     vector. The DH10B E. coli host was transformed with the
                     ligation mixture by the standard chemical method. The
                     average insert size is about 4.0 kb. The library was
                     constructed by Yulan Piao."
ORIGIN

  Query Match             17.3%;  Score 33.2;  DB 7;  Length 467;
  Best Local Similarity   52.1%;  Pred. No. 19;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
              |||||   ||| ||||  ||  | |  |     ||  |  |   |  | |  | ||||||
Db        273 AAGTGTGGTGATAAGCCAAACATGGATTCCACAGTGAAACTAGGAGCCGTGGGTGGAAAT 332

Qy         80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGAT 139
               ||| | ||   |||  || ||| |    ||| | | | | || | | | |    | |
Db        333 GGATTTAAAGTTCCTCTTACAACACCTCATTTGGAAAAGAGATACTCGATTCAGTTCGGG 392

Qy        140 GACAACATGGACTAAGCAAAAG 161
               ||||    ||  ||| | |||
Db        393 AACAATGAAGAAGAAGAAGAAG 414
```

```
RESULT 14
CK391130
LOCUS       CK391130                 493 bp    mRNA    linear   EST 29-DEC-2003
DEFINITION  K0827F01-5 NIA Mouse 8.5-dpc Whole Embryo cDNA Library (Long) Mus
            musculus cDNA clone NIA:K0827F01 IMAGE:30080988 5', mRNA sequence.
ACCESSION   CK391130
VERSION     CK391130.1  GI:40381649
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
  TITLE     Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
  JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
  MEDLINE   21429098
   PUBMED   11544199
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: K0827  row: F  column: 01
            Seq primer: M13 Reverse
            High quality sequence stop: 493
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..493
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:K0827F01-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:K0827F01 IMAGE:30080988"
                     /tissue_type="whole embryo including extraembryonic
                     tissues at 8.5-days postcoitum"
                     /dev_stage="8.5-days postcoitum"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 8.5-dpc Whole Embryo cDNA Library
                     (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     extracted from a pool of 13 embryos at 8.5-days
                     postcoitum. Double-stranded cDNAs were synthesized with an
                     Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     9.1 ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.5 kb. The library was constructed
                     by Yulan Piao (NIA)."
ORIGIN

  Query Match             17.3%;  Score 33.2;  DB 7;  Length 493;
  Best Local Similarity   52.1%;  Pred. No. 19;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
```

```
              |||||   ||| ||||  ||  | |  |     ||  |  |   |  | |  | ||||||
Db        273 AAGTGTGGTGATAAGCCAAACATGGATTCCACAGTGAAACTAGGAGCCGTGGGTGGAAAT 332

Qy         80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGAT 139
               ||| | ||   |||  || ||| |    ||| | | | | || | | | |    | |
Db        333 GGATTTAAAGTTCCTCTTACAACACCTCATTTGGAAAAGAGATACTCGATTCAGTTCGGG 392

Qy        140 GACAACATGGACTAAGCAAAAG 161
               ||||    ||  ||| | |||
Db        393 AACAATGAAGAAGAAGAAGAAG 414


RESULT 15
CD552026
LOCUS       CD552026                 501 bp    mRNA    linear   EST 11-JUN-2003
DEFINITION  B0337C08-5 NIA Mouse E9.5 Whole Embryo cDNA Library (Long) Mus
            musculus cDNA clone NIA:B0337C08 IMAGE:30432607 5', mRNA sequence.
ACCESSION   CD552026
VERSION     CD552026.1  GI:31599757
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
  TITLE     Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
  JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
  MEDLINE   21429098
   PUBMED   11544199
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0337  row: C  column: 08
            Seq primer: M13 Reverse
            High quality sequence stop: 501
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..501
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:B0337C08-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0337C08 IMAGE:30432607"
                     /tissue_type="E9.5 whole embryo"
                     /dev_stage="whole embryo including extraembryonic tissues
                     at 9.5-days postcoitum"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse E9.5 Whole Embryo cDNA Library
                     (Long)"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
                     Site_2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library (Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]).Total
                     RNAs were extracted from a pool of 16 embryos at 9.5-days
                     postcoitum. Double-stranded cDNAs were synthesized with an
                     Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3']
                     from6.1 ug of total RNA, treated with T4 DNA polymerase,
                     and purified by ethanol-precipitation. The cDNAs were
                     ligated to Lone-linker LL-Sal4, purified by
                     phenol/chloroform, and separated from free linkers by
                     Centricon 100. Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S. The products were purified
                     by phenol/chloroform and Centricon 100. The cDNAs were
```

```
                  digested with SalI and NotI enzymes and cloned into
                  SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E.
                  coli host was transformed with the ligation mixture by the
                  standard chemical method. The average insert size is about
                  3.0kb. The library was constructed by Yulan Piao."
ORIGIN

  Query Match             17.3%;  Score 33.2;  DB 6;  Length 501;
  Best Local Similarity   52.1%;  Pred. No. 20;
  Matches   74;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
               |||||   ||| ||||  ||  | |  |     ||  |  |   |  | |  | ||||||
Db         273 AAGTGTGGTGATAAGCCAAACATGGATTCCACAGTGAAACTAGGAGCCGTGGGTGGAAAT 332

Qy          80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGAT 139
                ||| | ||   |||  || ||| |    ||| | | | | || | | | |    | |
Db         333 GGATTTAAAGTTCCTCTTACAACACCTCATTTGGAAAAGAGATACTCGATTCAGTTCGGG 392

Qy         140 GACAACATGGACTAAGCAAAAG 161
                ||||    ||  ||| | |||
Db         393 AACAATGAAGAAGAAGAAGAAG 414

Search completed: August  6, 2005, 01:30:34
Job time : 1954.31 secs
```